



IJCSIS Vol. 14 No. 4, April 2016
ISSN 1947-5500

International Journal of Computer Science & Information Security

© IJCSIS PUBLICATION 2016
Pennsylvania, USA

Indexed and technically co-sponsored by :



AUTHOR SERIES



IJCSIS

ISSN (online): 1947-5500

Please consider to contribute to and/or forward to the appropriate groups the following opportunity to submit and publish original scientific results.

CALL FOR PAPERS

International Journal of Computer Science and Information Security (IJCSIS) January-December 2016 Issues

The topics suggested by this issue can be discussed in term of concepts, surveys, state of the art, research, standards, implementations, running experiments, applications, and industrial case studies. Authors are invited to submit complete unpublished papers, which are not under review in any other conference or journal in the following, but not limited to, topic areas.

See authors guide for manuscript preparation and submission guidelines.

Indexed by Google Scholar, DBLP, CiteSeerX, Directory for Open Access Journal (DOAJ), Bielefeld Academic Search Engine (BASE), SCIRUS, Scopus Database, Cornell University Library, ScientificCommons, ProQuest, EBSCO and more.

Deadline: see web site

Notification: see web site

Revision: see web site

Publication: see web site

Context-aware systems
Networking technologies
Security in network, systems, and applications
Evolutionary computation
Industrial systems
Evolutionary computation
Autonomic and autonomous systems
Bio-technologies
Knowledge data systems
Mobile and distance education
Intelligent techniques, logics and systems
Knowledge processing
Information technologies
Internet and web technologies
Digital information processing
Cognitive science and knowledge

Agent-based systems
Mobility and multimedia systems
Systems performance
Networking and telecommunications
Software development and deployment
Knowledge virtualization
Systems and networks on the chip
Knowledge for global defense
Information Systems [IS]
IPv6 Today - Technology and deployment
Modeling
Software Engineering
Optimization
Complexity
Natural Language Processing
Speech Synthesis
Data Mining

For more topics, please see web site <https://sites.google.com/site/ijcsis/>

arXiv.org Google scholar

SCIRUS
search engine for science

ScientificCommons

Scribd

docstoc
find and share professional documents

BASE
Bielefeld Academic Search Engine

CiteSeer^x beta

dblp.uni-trier.de
Computer Science
Bibliography

DOAJ
DIRECTORY OF
OPEN ACCESS
JOURNALS



ProQuest

For more information, please visit the journal website (<https://sites.google.com/site/ijcsis/>)

Editorial

Message from Editorial Board

It is our great pleasure to present the **April 2016 issue** (Volume 14 Number 4) of the **International Journal of Computer Science and Information Security (IJCSIS)**. High quality survey and review articles are proposed from experts in the field, promoting insight and understanding of the state of the art, and trends in computer science and technology. It especially provides a platform for high-caliber researchers, practitioners and PhD/Doctoral graduates to publish completed work and latest development in active research areas. According to Google Scholar, up to now papers published in IJCSIS have been cited over 5908 times and the number is quickly increasing. This statistics shows that IJCSIS has established the first step to be an international and prestigious journal in the field of Computer Science and Information Security. There have been many improvements to the processing of papers; we have also witnessed a significant growth in interest through a higher number of submissions as well as through the breadth and quality of those submissions. IJCSIS is indexed in major academic/scientific databases and repositories: Google Scholar, Thomson Reuters, ArXiv, CiteSeerX, Cornell's University Library, Ei Compendex, ISI Scopus, DBLP, DOAJ, ProQuest, ResearchGate, Academia.edu and EBSCO among others.

On behalf of IJCSIS community and the sponsors, we congratulate the authors and thank the reviewers for their outstanding efforts to review and recommend high quality papers for publication. In particular, we would like to thank the international academia and researchers for continued support by citing papers published in IJCSIS. Without their sustained and unselfish commitments, IJCSIS would not have achieved its current premier status.

"We support researchers to succeed by providing high visibility & impact value, prestige and excellence in research publication." For further questions or other suggestions please do not hesitate to contact us at ijcsiseditor@gmail.com.

A complete list of journals can be found at:
<http://sites.google.com/site/ijcsis/>

IJCSIS Vol. 14, No. 4, April 2016 Edition

ISSN 1947-5500 © IJCSIS, USA.

Journal Indexed by (among others):



Open Access This Journal is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source.



Bibliographic Information

ISSN: 1947-5500

Monthly publication (Regular Special Issues)

Commenced Publication since May 2009

Editorial / Paper Submissions:

IJCSIS Managing Editor

[\(ijcsiseditor@gmail.com\)](mailto:ijcsiseditor@gmail.com)

Pennsylvania, USA

Tel: +1 412 390 5159

IJCSIS EDITORIAL BOARD

Editorial Board Members	Guest Editors / Associate Editors
Dr. Shimon K. Modi [Profile] Director of Research BSPA Labs, Purdue University, USA	Dr. Riktesh Srivastava [Profile] Associate Professor, Information Systems, Skyline University College, Sharjah, PO 1797, UAE
Professor Ying Yang , PhD. [Profile] Computer Science Department, Yale University, USA	Dr. Jianguo Ding [Profile] Norwegian University of Science and Technology (NTNU), Norway
Professor Hamid Reza Naji , PhD. [Profile] Department of Computer Engineering, Shahid Beheshti University, Tehran, Iran	Dr. Naseer Alquraishi [Profile] University of Wasit, Iraq
Professor Yong Li , PhD. [Profile] School of Electronic and Information Engineering, Beijing Jiaotong University, P. R. China	Dr. Kai Cong [Profile] Intel Corporation, & Computer Science Department, Portland State University, USA
Professor Mokhtar Beldjehem , PhD. [Profile] Sainte-Anne University, Halifax, NS, Canada	Dr. Omar A. Alzubi [Profile] Prince Abdullah Bin Ghazi Faculty of Information Technology Al-Balqa Applied University (BAU), Jordan
Professor Yousef Farhaoui , PhD. Department of Computer Science, Moulay Ismail University, Morocco	Dr. Jorge A. Ruiz-Vanoye [Profile] Universidad Autónoma del Estado de Morelos, Mexico
Dr. Alex Pappachen James [Profile] Queensland Micro-nanotechnology center, Griffith University, Australia	Prof. Ning Xu , Wuhan University of Technology, China
Professor Sanjay Jasola [Profile] Dean, School of Information and Communication Technology, Gautam Buddha University	Dr. Bilal Alatas [Profile] Department of Software Engineering, Firat University, Turkey
Dr. Siddhivinayak Kulkarni [Profile] University of Ballarat, Ballarat, Victoria, Australia	Dr. Ioannis V. Koskosas , University of Western Macedonia, Greece
Dr. Reza Ebrahimi Atani [Profile] University of Guilan, Iran	Dr Venu Kuthadi [Profile] University of Johannesburg, Johannesburg, RSA
Dr. Umar Ruhi [Profile] University of Ottawa, Canada	Dr. Zhihan Lv [Profile] Chinese Academy of Science, China
Dr. Vahid Esmaeizadeh [Profile] Iran University of Science and Technology	Prof. Ghulam Qasim [Profile] University of Engineering and Technology, Peshawar, Pakistan
Dr. Jiliang Zhang [Profile] Northeastern University, China	Prof. Dr. Maqbool Uddin Shaikh [Profile] Preston University, Islamabad, Pakistan
Dr. Jacek M. Czerniak [Profile] Casimir the Great University in Bydgoszcz, Poland	Dr. Musa PEKER [Profile] Faculty of Technology, Mugla Sitki Kocman University, Turkey
	Dr. Wencan Luo [Profile] University of Pittsburgh, US

TABLE OF CONTENTS

1. PaperID 31031698: Flexible Hardware Implementation of Hyperelliptic Curves Cryptosystem (pp. 1-7)

*Anissa Sghaier #, Chiraz Massoud #, Medien Zeghid *, Mohsen Machhout #*

Faculty of Sciences, University of Monastir, EμE Lab Avenue of the environment 5019 Monastir

** Higher Institute of Applied Sciences and Technology Taffala City, 4003 Sousse Tunisie*

Abstract — Due to the resolution of key distribution problem, asymmetric scheme become the most popular cryptographic technique compared to symmetric scheme. One of the well-known asymmetric encryption algorithms are (Hyper)Elliptic Curve Cryptosystem (ECC and HECC). They provide equal security levels compared with the RSA algorithm with shorter operand size. Although, the HECC outperform the ECC due to its shorter operand size. The objective of this paper is to present an efficient hardware architecture using Cantor's method, to implement a new way of explicit formula over genus curve 2, and analyze the performance of the two implementations. HECC cryptosystem was implemented over GF(283) on XC5V240 FPGA, it takes about 5086 slices, and it runs at 175 MHz in 0.287 ms.

Keywords: ECC, HECC, hardware implementation, Cantor's method, explicit formula

2. PaperID 310316100: Phishing Identification Using a Novel Non-Rule Neuro-Fuzzy Model (pp. 8-14)

Luong Anh Tuan Nguyen, Faculty of Information Technology, Ho Chi Minh City University of Transport, Ho Chi Minh City, Vietnam

Huu Khuong Nguyen, Faculty of Information Technology, Ho Chi Minh City University of Transport, Ho Chi Minh City, Vietnam

Abstract — This paper presents a novel approach to overcome the difficulty and complexity in identifying phishing sites. Neural networks and fuzzy systems can be combined to join its advantages and to cure its individual illness. This paper proposed a new neuro-fuzzy model without using rule sets for phishing identification. Specifically, the proposed technique calculates the value of heuristics from membership functions. Then, the weights are trained by neural network. The proposed technique is evaluated with the datasets of 22,000 phishing sites and 10,000 legitimate sites. The results show that the proposed technique can identify with an accuracy identification rate of above 99%.

Keywords — Phishing; Fuzzy; Neural Network; Neuro-Fuzzy

3. PaperID 31031614: A Predictable Markov Based Cache Replacement Scheme in Mobile Environments (pp. 15-26)

Ahmed. A. A. Gad-ElRab, Faculty of Science, Al-Azhar University, Cairo, Egypt

Kamal A. ElDahshan, Faculty of Science, Al-Azhar University, Cairo, Egypt

Ahmed Sobhi, (PhD Student) Faculty of Science, Al-Azhar University, Cairo, Egypt

Abstract — Mobile Location-dependent services are popular services that the mobile environments support. Data caching is an effective technique that plays an important role in improving these services. In mobile environments, due to the limited cache size of mobile devices, the cache replacement which is finding a suitable subset of items for eviction from cache becomes important. Most of the existing cache replacement schemes use the cost functions in the replacement operation. In this paper we propose a Predictable Markov based Cache Replacement (PMCR) scheme for Mobile Environments. The proposed scheme uses a markov model with cost function in the replacement operation. The key idea of the markov model is the prediction of future client locations by giving us the weight of visiting each location whose data is cached. Simulation results show that our approach improves the system performance compared to the existing schemes.

Keywords - Mobile Location-dependent services; Data dissemination; Cache replacement; Predicted region; Markov model; PMCR.

4. PaperID 31031619: Developing an Intelligent System for Crowd Density Estimation (pp. 27-31)

*Dr. Ali Salem Ali Bin-Sama, Dr. Salem Saleh Ahmed Alamri
Department of Engineering Geology, Oil & Minerals Faculty, Aden University, Aden, Yemen*

Abstract — Crowd density estimation models are important for monitoring people behaviors in a crowd. In this paper a development of an intelligent system is introduced to achieve the goal of density estimation. Mainly, the proposed system consist of Gabor features texture pattern extraction and convolutional neural network for pattern classification. To assess the performance of the developed, a number of public benchmark images are used such as LIBRARY Dataset, QUT Dataset, and Fudan Pedestrian Dataset.

Keyword: Crowd Density Estimation, Gabor filters, Convolutional neural network, Texture Image.

5. PaperID 31031641: A Security Scheme for Providing AIC Triad in Mobile Cloud Computing (pp. 32-36)

*Isra Sitan Al-Qasrawi
Department of Information Technology, Al-balqa' Applied University, AL-Huson University College, Irbid, Jordan*

Abstract — As mobile devices like smart phones and tablets continue to grow, the requirement of cloud computing in mobile devices continue to grow too, and becomes an important service to provide users the ability to manage files and data remotely, which gave birth of Mobile Cloud Computing (MCC). As a result, new web-based threats and attacks will continue to increase in number. The most important issues must be covered to provide users reliable and secure services of mobile cloud computing are: Availability, Integrity, and Confidentiality. In this paper, (i) the concepts of cloud computing and mobile computing are discussed, the challenges that face each one of them, the meaning of mobile cloud computing, the challenges of MCC. (ii) Different mechanisms to store data in secure manner are explored. (iii) Propose a new scheme to secure the data storage in Mobile Cloud Computing without exposing the data content to the cloud service providers to protect mobile users' privacy. This scheme provides the security AIC triad concepts (Availability, Integrity, and Confidentiality) for data by applying a number of operations.

Keywords - cloud computing; mobile computing; mobile cloud computing; security; data storage; mobile user.

6. PaperID 31031668: SimCT: A Measure of Semantic Similarity Adapted to Hierarchies of Concepts (pp. 37-44)

*Coulibaly Kpinna Tiekoura, National Polytechnic Institute, Department of Mathematics and Computer Science, Abidjan, Ivory Coast
Brou Konan Marcellin, National polytechnic Institute, Department of Mathematics and Computers Science, Yamoussoukro, Ivory Coast
Achiepo Odilon, National polytechnic Institute, Abidjan, Ivory Coast
Babri Michel, National Polytechnic Institute, Abidjan, Ivory Coast
Aka Boko, University of Nangui Abrogoua, Abidjan, Ivory Coast*

Abstract — The Calculating of the similarity between data is a key problem in several disciplines such as machine learning, information retrieval (IR) and data analysis. In some areas such as social resilience, the similarity measures can be used to find the similarities between traumatized individuals or resilience's dimensions. In this paper, we propose a measure of semantic similarity used in many applications including clustering and information retrieval. It relies on a knowledge base represented as a hierarchy of concepts (ontology, graph, taxonomy). Its uniqueness with respect to previous proposals is the difference between the indices of similarity that it establishes between brothers concepts located at the same hierarchical level and having the same direct ancestor. In addition, our semantic similarity

measure provides better modularity in clustering compared with Wu and Palmer's similarity measure and Proxymeneas [3].

Keywords - clustering, hierarchical tree, resilience, semantic similarity measure.

7. PaperID 31031672: An Algorithm (COLMSTD) for Detection of Defects on Rail and Profile Surfaces (pp. 45-50)

İlhami Muharrem Orak, Faculty of Engineering, Karabük University, Karabük, Turkey (78000)
Ahmet Çelik, Tavşanlı Vocational School, Dumlupınar University, Kütahya, Turkey

Abstract — Rail or profile products are used in many fields today. The rolling process is the most important production phase of the rail and the profile product. However, undesirable defects in the surface of the product during the rolling process can occur. Identifying these defects quickly by an intelligent system using image processing algorithms will provide a major contribution in terms of time and labor. For the detection of the regions, objects and shapes on the image, several algorithms were used. In this study, we introduce a Standard Deviation based algorithm (COLMSTD) by using the pixel color values. In order to evaluate the performance of the algorithm, the result of the COLMSTD algorithm is compared with the results of Hough Transform, MSER, DFT, Watershed, Blob Detection algorithms. In this study, it was seen that each algorithm has different capability in some extent to identify the surface defects in rail or profile. However, COLMSTD algorithm achieve more accurate and successful results than the other algorithms.

Keywords - Computer vision; Image processing; Manufacturing systems; Defect detection; Hot rolling; Rail; Profile.

8. PaperID 310316121: Investigating the Opportunities of Using Mobile Learning by Young Children in Bulgaria (pp. 51-55)

*Radoslava Krалева #, Aleksandar Stoimenovski #, Dafina Kostadinova *, Velin Krалев #*
Department of Informatics, South West University "Neofit Rilski", Blagoevgrad, Bulgaria
** Department of Germanic and Romance Studies, South West University "Neofit Rilski", Blagoevgrad, Bulgaria*

Abstract – This paper provides an analysis of literature related to the use of mobile devices in teaching young children. For this purpose, the most popular mobile operating systems in Bulgaria are considered and the functionality of the existing mobile applications with Bulgarian interface is discussed. The results of a survey of parents' views regarding the mobile devices as a learning tool are presented and the ensuing conclusions are provided.

Keywords – Mobile learning, Mobile learning application, Analysis of the parents' opinion

9. PaperID 31031638: Conducting Multi-class Security Metrics from Enterprise Architect Class Diagram (pp. 56-61)

Osamah S. Mohammed, Dept. of Software Engineering, College of Computer Sc. & Math, University of Mosul. Mosul, Iraq.
Dujan B. Taha, Dept. of Software Engineering, College of Computer Sc. & Math, University of Mosul. Mosul, Iraq

Abstract — Developers often neglect security until the end of developing the software just after coding, and any change in the code with respect to security may lead to change in the software code, this consumes time and cost depending on the software size. Applying security on a software late in its SDLC may result in many security flaws, some of them can involve serious architectural issues. Applying security metrics on design phase can reveal the security level and fix vulnerabilities of a software earlier in the project. In this work, security metrics has been discussed, and conducting these metrics from Enterprise Architect class diagram using a proposed CASE tool.

Keywords - Software Engineering; Security metrics; Enterprise architect; Class diagram; SDLC; Design phase

10. PaperID 31031639: Data Traffic Optimization in Different Backoff Algorithms for IEEE 802.15.4/Zigbee Networks (pp. 62-66)

Muneer Bani Yassein, Maged Refat Fakirah

Faculty of Computer and Information Technology, Jordan University of Science and Technology, Irbid, Jordan

Qusai Abuein, Mohammed Shatnawi, Laith Bani Yaseen

Jordan University of Science and Technology Irbid, Jordan

Abstract — Zigbee/IEEE 802.15.4 is a short range wireless communication standard designed for home monitoring, health care, and industrial applications. In this paper, the impact of data traffic load and two data traffic types, namely, Constant Bit Rate (CBR) and Variable Bit Rate (VBR) are studied by considering Binary Exponential Backoff Algorithm (BEB), Liner Backoff Algorithm and Fibonacci Backoff Algorithm (FIB). The efficiency of these algorithms is extensively evaluated by modifying the number of CBR or VBR packets sent from the nodes to the PAN coordinator. The obtained results demonstrate that using the VBR data traffic increases the throughput and decreases the end to end delay, while adopting the CBR data traffic decreases the total energy consumption of a small scale network.

Keywords—IEEE 802.15.4/ZigBee; backoff ; BEB; Linear; FIB; data traffic load; VBR; CBR

11. PaperID 31031653: A Novel Polygon Cipher Technique using Hybrid Key Scheme (pp. 67-71)

Shadi R. Masadeh, Faculty of Information Technology, Isra University, Amman, Jordan

Hamza A. A. Al_Sewadi, King Hussein Faculty of Computing, Prince Sumaya for Technology, Amman, Jordan

Abstract — Due to the narrow key space and frequency analysis weakness, classical cipher techniques are not suitable for most today's information communication. On the other hand, modern standardize ciphers are far more secure and widely used for such communication. However, they are so complicated in implementation and may not be suitable for less sophisticated applications. This paper suggests a novel symmetric cipher method based on polygon scheme that shows superior security as compared with classical methods by having wide key space and strength against frequency analysis attack and yet it is simpler than modern ciphers.

Keywords- information security, encryption/decryption, secret key, symmetric cryptography, asymmetric key implementation.

12. PaperID 31031659: An Efficient Method to Diagnose the Treatment of Breast Cancer using Multi-Classifiers (pp. 72-80)

J. Umamaheswari, Computer Science dept. Majmaah University, Al- Majmaah, Saudi Arabia

Jabeen Sultana, Ruhi Fatima, Computer Science dept. Majmaah University, Al- Majmaah, Saudi Arabia

Abstract — Knowledge discovery in the form of rule extraction proposed to extract rules from classification datasets by giving data set to Decision Trees (DT), NBTREE, KNN and 10-fold Cross Validation performed, resulting the tree or a model from which rules are extracted and measured on different parameters taken from root node to leaf node.

Keywords - Transparent; Opaque; Knowledge discovery; rule extraction

13. PaperID 31031607: A Study on Optimizing the Efficiency of Location Aided Routing (LAR) Protocol (pp. 81-86)

Priyanka Kehar, Department of Computer Science, Lovely Professional University, Punjab, India

Pushpendra Kumar Pateriya, Lovely Faculty of Technology and Sciences, Lovely Professional University, Phagwara, India

Abstract -The improvised network is an arrangement less network consisting of portable nodes. VANETs is the recently developed technique to achieve traffic safety and efficiency through inter vehicle communication, where routing protocol plays a vital role. Inefficient path establishment and network congestion both bring the severe degradation in network throughput and performance. Routing throughput and enactment is largely reliant on the stability and availability of the wireless link which makes it a very pivotal factor, that can't be ignored in order to obtain proper performance and throughput measurement in vehicular improvised network. As vehicle nodes have higher mobility due to which some prediction based techniques were proposed in previous times for path establishment. Among the proposed prediction based techniques, location aided routing protocol influence real time vehicular information to generate path between source and destination, with high possibility of network connectivity among them. The main feature of optimized LAR is: minimize the delay, minimize the fuel consumption, and maximize the throughput.

Keywords - Road Side Unit (RSU); Location Aided Protocol (LAR); Internet Service Provider (ISP); Intelligent Transport Service (ITS).

14. PaperID 31031611: Analyzing and Processing Data Faster Based on Balanced Partitioning (pp. 87-92)

Annie P. Kurian, Dept. of Computer Science & Engg., Velammal Engg. College, Chennai, India
Prof. Dr. V. Jeyabalaraja, Dept. of Computer Science & Engg., Velammal Engg. College, Chennai, India

Abstract — Big data has become a well-known buzzword to the public at large which handles enormous amount of data i.e., in terabyte to zeta byte. Processing and analyzing such huge amount of data is not possible with traditional and conventional environments. The existing system approaches for range partition queries are deficient to rapidly provide definite results in big data. In this paper, we propose a agile approach to range- aggregate queries in big data documents/table using balanced partitioning. This approach first divides the big data into independent partition with balanced partitioning, and then it generates a local estimation sketch for each partition. When a RA-query request arrives, the system quickly fetches and obtains the result directly by compiling local estimation from all partitions. The balanced partitioning avoids the overall scan of the data in order to provide the result. Big data ecosystem like HIVE and Impala is used to handle the structured data and uses the balanced partitioning to provide fast and accurate output. Partitioning provides maintenance, availability and improvised query performance to the users. It reduces the time complexity, i.e., $O(1)$ time complexity for data updates. The overall performance of the dataset produced would be efficient, fault-tolerant, accurate and fast.

Keywords – range aggregate, big data, HIVE, Impala, partition, map reduce, HDFS.

15. PaperID 31031613: ICT Convergence in Internet of Things – The Birth of Smart Factories (pp. 93)

Mahmood Adnan, Hushairi Zen
Faculty of Engineering, Universiti Malaysia Sarawak

Abstract – Over the past decade, most factories across developed parts of the world employ a varying amount of the manufacturing technologies including autonomous robots, RFID (radio frequency identification) technology, NCs (numerically controlled machines), wireless sensor networks embedded with specialized computerized softwares for sophisticated product designs, engineering analysis, and remote control of machinery, etc. The ultimate aim of these all dramatic developments in manufacturing sector is thus to achieve aspects such as shorter innovation / product life cycles and raising overall productivity via efficiently handling complex interactions among the various stages (functions, departments) of a production line. The notion, Factory of the Future, is an unpredictable heaven of efficaciousness, wherein, issues such as the flaws and downtime would be issues of the long forgotten age. This technical note thus provides an overview of this awesome revolution waiting to be soon realized in the manufacturing sector.

16. PaperID 31031626: IEEE 802.11ac vs IEEE 802.11n: Throughput Comparison in Multiple Indoor Environments (pp. 94-101)

*Zawar Shah (a), Ashutosh A Kolhe (a), Omer Mohsin Mubarak (b)
(a) Whitireia Community Polytechnic, Auckland, New Zealand.
(b) Iqra University, Islamabad, Pakistan*

Abstract — IEEE 802.11ac is a fifth generation WiFi standard that has many advanced features than the current widely used IEEE 802.11n. In this paper, we perform experiments in two real indoor environments (that possess interference and have different multipath characteristics) to quantify the gain in average throughput provided by IEEE 802.11ac compared to IEEE 802.11n. Our experimental results show that in an environment with less multipath effect, IEEE 802.11ac provides 51% and 126% gain compared to IEEE 802.11n at a distance of 5m and 18.5m from the wireless router, respectively. Similarly, in an environment with high multipath effect, IEEE 802.11ac provides gain of 21% and 32% compared to IEEE 802.11n at a distance of 1m and 18.5m from the wireless router, respectively. We conclude that IEEE 802.11ac can effectively handle interference caused by other IEEE 802.11n (5GHz) sources and provides higher throughput than IEEE 802.11n.

Keywords: IEEE 802.11ac, IEEE 802.11n, Throughput, MIMO.

17. PaperID 31031651: Implementing Navigational Aspect of Specific Testing Process Model (pp. 102-111)

*Garima Singh, Dept. of Computer Science and Engineering, JECRC University, Jaipur, Rajasthan, India
Manju Kaushik, Associate Professor, Dept. of Computer Science and Engineering, JECRC University, Jaipur, Rajasthan, India*

Abstract - Navigational modeling of web application and testing the navigational aspect of the web application is as important as the content displayed and security of application to maintain the quality and user satisfaction. Test paths are generated through the navigation model which is derived from the activity diagram. The objective of this paper is to implement navigational aspect of web application through a model.

Keywords - Specific Testing Process Model, Web application modelling, web application navigational testing

18. PaperID 31031667: Comparative Analysis of LEACH and V-LEACH Protocols in Wireless Sensor Networks (pp. 112-119)

*Layla Aziz (*1), Said Raghay (1), Abdellah Jamali (2), and Hanane Aznaoui (1)
(1) Laboratory(LAMAI),Cadi Ayyad University, Marrakech, Morocco
(2) Laboratory (RI2M), Hassan 1st University, Berrchid, Morocco*

Abstract — In the past few years, the research community is strongly attracted to wireless sensor networks (WSNs). Sensor node is generally driven by an irreplaceable battery which limits its energy supply. A number of new methods and strategies have been proposed to reduce energy consumption in WSNs. LEACH (Low Energy Adaptive Clustering Hierarchy) protocol is a well-known approach using the Clustering mechanism to minimize the energy consumption and improve the lifetime of WSN . In this work, we describe various clustering algorithms and a comparative analysis of LEACH protocol with its improved version V-LEACH using NS2 simulator.

Index Terms— CLUSTERING, LEACH, NS2, V-LEACH, WSN

19. PaperID 31031670: Slow Wave-IDC Loaded High Bandwidth Microstrip Antenna Operates For Multi Band Applications (pp. 120-125)

Brajlata Chauhan, Uttarakhand Technical University, Dehradun UK, India
Sandip Vijay, Deptt. of Electronics & Communication Engg. ICFAI Univ. Dehradun UK, India
S C Gupta, Department of Electronics & Communication Engineering, DIT Dehradun UK, India

Abstract — A slow wave structure as inter-digital capacitor (IDC) is incorporated in micro-strip patch to obtain Miniaturized and high band width antenna specially for WLAN, X & Ku –bands. The antennas are loaded with IDC to slow down the guided wave to increase Gain - Bandwidth product. The simulated antennas offered gain of 6.47dB, directivity of 6.47dB and radiated power of 0.001066 watt (antenna2). This paper presents increased bandwidth to 55.33% by inserting a slot on the above patch offered nominal change in gain of 5.8852 and the loaded slot antenna produce directivity of 7.38832dB and radiated power of 0.0299368 watt (antenna 3) in the range of VSWR is less than 1.5.

Keywords- *Slow wave structure; inter-digital capacitor (IDC); Gain band width product; multi band micro-strip patch antenna; rectangular slot; equivalent circuit.*

20. PaperID 31031673: An Efficient Anti-noise Fast FCM Clustering for Glioblastoma Multiforme Tumor Segmentation (pp. 126-133)

B. Srinivasa Rao, ANUCET, Acharya Nagarjuna University, Guntur-522510, Andhra Pradesh, India.
Dr. E. Sreenivas Reddy, Professor, ANUCET, ANUCET, Acharya Nagarjuna University, Guntur-522510, Andhra Pradesh, India

Abstract -- Image segmentation plays an important role in medical image processing. Magnetic Resonance Imaging (MRI) is primary diagnostic technique to do image segmentation. Clustering is an unsupervised learning method of segmentation. The conventional FCM algorithm is sensitive to noise, suffers from the computation time overhead and is very sensitive to cluster center initialization. In order to overcome this problem, a new method called Anti-Noise Fast Fuzzy C-Means (AN-FFCM) clustering algorithm for segmentation of Glioblastoma Multiforme tumor segmentation is proposed. The proposed algorithm is able to minimize the effects of impulse noise by incorporating noise detection stage to the clustering algorithm during the segmentation process without degrading the fine details of the image. This method also improves the performance of the FCM algorithm by finding the initial cluster centroids based on histogram analysis, reducing the number of iterations for segmentation of noisy images. The advantages of the proposed method are: (1) Minimizes the effect of impulse noise during segmentation, (2) Minimum number of iterations to segment the image. The performance of the proposed method is tested on BRATS data set. Experimental results show that the proposed algorithms are superior in preserving image details and segmentation accuracy while maintaining a low computational complexity.

Index Terms: *Glioblastoma Multiforme(GBM), image segmentation, Histogram, salt-and-pepper noise, Fuzzy c-means, Medical Image processing.*

21. PaperID 31031679: Ear Classifying in Profile images Based on Curves and Shape (pp. 134-137)

Mohammad Mahmoudi, Department of Computer Science and Engineering, Khoy branch, Islamic Azad University Khoy, Iran
Ali Habiboghli, Department of Computer Science and Engineering, Khoy branch, Islamic Azad University Khoy, Iran

Abstract — In this research we are going to classify ears based on their appearance. For this aim, region of ear in profile image should be extracted. Then by using margins surrounding around the ear and the center of ear would be obtained by the proposed method. Finally by determining appropriate threshold the ears were classified based on their shapes. The database used in this article is CVL. Simulating and classifying of this article have acceptable accuracy 83.6%.

Keywords -- *Classification, Ear Recognition; Image Processing; Profile Images*

22. PaperID 31031685: Color Coding Based Detection and Prevention Mechanism for Wormhole Attack in MANET (pp. 138-144)

Harsh Bansal, Lovely Professional University, Phagwara, Punjab, India
Gurpreet Singh, Lovely Professional University, Phagwara, Punjab, India

Abstract — MANET is infrastructure-less, lacks centralized monitoring and has dynamic changing network topology. The high usage of MANET demands more security and confidentiality and integrity of the data communicated through network. Security has turned out to be a major concern so as to provide non-endangered communication between mobile nodes in an unfriendly environment of MANET, which poses a number of trivial challenges to security design. The wormhole attack is one of the most threatening and hazardous attacks. In this paper we have classified the well-known countermeasures against wormhole attack in the network according to detection and prevention techniques based on hop counts and delay, protocol modification, trust and reputation. The projected technique to be used for detection of wormhole attack using trust based mechanism, neighbor monitoring concept and credits based mechanism will help to detect and isolate the malicious nodes hence enabling the formation of trusted network.

Keywords— MANET, Intrusion Detection, Wormhole Attack, Secure Routing, Network Security.

23. PaperID 31031687: Pragmatic Analysis Based Document Summarization (pp. 145-149)

*Ms. Bhakti A. Mukhedkar #, Mrs. D. Y. Sakhare #, Mr. Raj Kumar **
Department of Electronics Engineering, MIT Academy of Engineering, Alandi, Pune, India
** Scientist DRDO, Pune.*

Abstract - Automatic Text summarization is the process of reducing a text document to create a summary that relates only important points of the original document. Now a day's huge information available so there is interest in automatic Text summarization. It's very hard for human being to manually summarize large documents of text. Hence we use Text Summarization techniques. Basically Text Summarization Techniques classified in two types 1. Abstraction 2. Extraction. In this Paper We Proposed Abstraction Type of Text Summarizations by using pragmatic analysis. This Summary being generated by Matlab and serially transmitted to PIC microcontroller and displayed on LCD.

Index Terms— POS Tagging, Text Summarization by pragmatic analysis.

24. PaperID 31031695: Mobility Aware Fault Tolerant Multipath Multicast Routing Protocol for MANET (pp. 150-158)

Channabasayya Mathad, Dept. of Electronics & Communication Engg, Kalpataru Institute of Technology, Tiptur Karnataka, India.
Paramesha, Dept. of Electronics & Communication Engg, Govt Engineering College, Hassan Karnatana, India.
D Srinivasa Rao, Dept. of Electronics & Communication Engg, Jawaharlal Nehru Technological University, Hyderabad Telangana, India

Abstract — In MANETs, due to the constant mobility of the nodes, the topology is ever changing. Hence, the selection of paths is crucial. So, it is always efficient to select more than one route to the destination, so that even if one path fails, there is always high possibility for the data to reach the destination. In MANETs, since the nodes keep on joining and leaving the network randomly, selecting paths that are less susceptible to turn out faulty is important. Since several disjoint paths are possible, multicasting is economical in MANETs. In this proposed scheme a multipath, multicast routing protocol which works efficiently by selecting route with higher lifetime and it also recovers the lost packets.

Keywords - Multipath, Multicast, Fault Tolerant, LinkLife Time, Hop Count.

25. PaperID 31031697: Fully Homomorphic Encryption: State of Art and Comparison (pp. 159-167)

Ahmed EL-YAHYAOU, Mohamed Dafir ELKETTANI

Information Security Research Team, CEDOC ST2I ENSIAS, Mohammed V University in Rabat, Rabat, Morocco

Abstract - Fully homomorphic encryption (FHE) is an alternative of cryptography that allows evaluating arbitrary functions on encrypted data without the need for decryption of ciphertexts. In this article we present the state of the art of fully homomorphic encryption schemes. In particular we present a classification of several existent FHE schemes followed by a comparison of performances and complexity of these cryptosystems. Finally we will give different possible axes of research in the conclusion.

Keywords: cryptosystem, fully homomorphic, cloud, bootstrappability, modulus reduction, key changing.

26. PaperID 310316106: Autoregressive Model Based Segmentation of Overlapped Region (pp. 168-174)

Vidyadevi G Biradar, Department of ISE, NMIT, Bangalore, India

H Sarojadevi, Department of CSE, NMAMIT, Bangalore, India

H C Nagaraj, Department of ECE, NMIT, Bangalore, India

Abstract — Overlapped fingerprints occur due to multiple impressions of fingerprints on the same object at same place. This is natural in uncontrolled environments, or they are the residual fingerprints left over on fingerprints scanner. Overlapped fingerprints need to be separated into individual fingerprints for recognition. Separation of overlapped fingerprints involves steps, segmentation of image regions, feature extraction and classification. State of the art algorithms for separation of overlapped fingerprints adopts region wise processing approach to feature extraction. Therefore segmentation of overlapped region is an essential step for robust feature extraction. This paper presents a new algorithm for segmentation of overlapped region using time series two dimensional Autoregressive (2D AR) model. AR model parameters are estimated using Least Squares (LS) method which ensures minimum mean square error. The performance of the algorithm is evaluated using a standard database of 100 overlapped fingerprints images. The results are compared with ground truth results and are found satisfactory. Segmentation accuracy achieved is between 80% to 90%.

Keywords- Segmentation, AR model, overlapped fingerprints, texture, separation.

27. PaperID 310316107: A Hybrid Hashing Security Algorithm for Data Storage on Cloud Computing (pp. 175-181)

Noha MM. AbdElnapi, Computer science department, Nahda University, Beni Suef, Egypt

Fatma A. Omara, Computer science department, Cairo University, Cairo, Egypt

Nahla F. Omran, Mathematics department, South Valley University, Qena, Egypt

Abstract — In today's modern IT everything is possible on the web by cloud computing, it allows us to create, configure, use and customize the applications, services, and storage online. The Cloud Computing is a kind of Internet-based computing, where shared data, information and resources are provided with computers and other devices on-demand. The Cloud Computing offers several advantages to the organizations such as scalability, low cost, and flexibility. In spite of these advantages, there is a major problem of cloud computing, which is the security of cloud storage. There are a lot of mechanisms that is used to realize the security of data in the cloud storage. Cryptography is the most used mechanism. The science of designing ciphers, block ciphers, stream ciphers and hash functions is called cryptography. Cryptographic techniques in the cloud must enable security services such as authorization, availability, confidentiality, integrity, and non-repudiation. To ensure these services of security, we propose an effective mechanism with a significant feature of the data. This paper is to show how to improve the security of the Cloud storage using the implementation of a hybrid encryption algorithm and hash functions. It proposes the implementation of two algorithms, Rivest-Shamir-Adleman (RSA) and Advanced Encryption Standard (AES) with a

secure hashing algorithm (SHA256) by using Netbeans IDE 8.0.2, JDK 1.7 tool and EyeOS2.5 as a cloud platform on ubuntu14.04.

Keywords— Cloud Computing, Security, Advanced Encryption Standard (AES), Rivest-Shamir-Adleman (RSA), Hybrid Algorithm, Hash functions, Secure Hash Algorithm (SHA256), Encryption, Cryptography, availability, confidentiality, integrity, authorization, and non-repudiation.

28. PaperID 310316108: 8-neighborhood Variant for a Better Container Code Extraction and Recognition (pp. 182-186)

*Wassim Al-Khawand, School of Engineering, Sciences and Technologies - University of Genoa, Italy
Seifedine Kadry, School of Engineering – American University of the Middle East, Kuwait
Riccardo Bozzo, Dept. of Electrical, Electronic, Telecommunications Engineering and Naval Architecture - University of Genoa, Italy
Khaled Smaili, Faculty of Sciences – Lebanese University, Lebanon*

Abstract — In this paper, we will present a new variant of the 8-neighborhood connectivity; our approach remedies the segmentation problem related to scratched container code digits. Our approach is highly suitable for real-time automatic container code recognition applications because it treats many special cases, its average response time is equal to 21 milliseconds, and it improves the container code extraction and recognition by 0.89%; due to our contribution in enhancing the segmentation phase, the container code extraction accuracy reached 98.7%.

Keywords— binary image, 8-neighborhood connectivity, segmentation, Container code.

29. PaperID 310316114: Notification System Based on Face Detection and Recognition: A Novel Approach (pp. 187-191)

*Ahmed AbdulQader Al-Bakeri, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia
Abdullah Ahmad Basuhail, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract — Nowadays, many applications implemented for face detection and recognition are used to achieve different types of projects, whether they are to be used for attendance systems in schools or for the check-in and check-out of employees in an organization. The purpose of this paper is to propose a new notification system using face detection and recognition to notify the house owner of visitors by using the SMTP to send an email containing the names and phone numbers of those visitors. In this system, the camera detects and recognizes the persons in front of the door and then sends their personal information to the host. The theoretical and practical aspects of this system are provided as follows.

Keywords- Face, Biometric, SMTP, Notification, Face recognition

30. PaperID 310316120: AndorEstimator: Android based Software Cost Estimation Application (pp. 192-202)

*Muhammad Zubair Asghar, Institute of Computing and Information Technology, Gomal University, D.I.Khan, Pakistan
Ammara Habib, Institute of Computing and Information Technology, Gomal University, D.I.Khan, Pakistan
Anam Habib, Institute of Computing and Information Technology, Gomal University, D.I.Khan, Pakistan
Syeda Rabail Zahra, Institute of Computing and Information Technology, Gomal University, D.I.Khan, Pakistan
Sadia Ismail, Institute of Computing and Information Technology, Gomal University, D.I.Khan, Pakistan*

Abstract — The main aim of the proposed system is to assist the software development team to estimate the cost, effort and maintenance of the project under development. Android-based platform, namely MIT App Inventor is used for

the development of application, which contains visual block programming language. The current study has following uniqueness of (1) Accuracy of results, (2) user friendly environment (3) no such application is available on android platform to the best of our knowledge. Questionnaire regarding CoCoMo model is developed and circulated by using objective qualitative method. Findings: The estimation module of our application is quite important with respect to facilitating the students of software engineering for performing CoCoMo-based cost estimation easily, and enabling the software developers for performing software cost estimation easily. The cost estimator based on CoCoMo model is developed on android platform however, to the best of our knowledge no such application is available. This system can be used by business and educational stakeholders, such as students, software developers, and business organizations.

Keywords — CoCoMo model; App Inventor; Cost estimation; Android

31. PaperID 310316129: Survey of Keystroke Dynamics as a Biometric for Static Authentication (pp. 203-207)

*Pranit Shinde, Dept. of Computer Engineering, Fr. Conceicao Rodrigues College of Engineering, Mumbai, India
Saideep Shetty, Dept. of Computer Engineering, Fr. Conceicao Rodrigues College of Engineering, Mumbai, India
Mahendra Mehra, Dept. of Computer Engineering, Fr. Conceicao Rodrigues College of Engineering, Mumbai, India*

Abstract — Keystroke Dynamics is the study of a user's typing pattern based on the various timing information obtained when a key is pressed and released. It comes under Behavioral Biometrics and has been a topic of interest for authenticating as well as identifying users based on their typing pattern. There have been numerous studies conducted on Keystroke Dynamics as a Biometrics with different data acquisition methods, user base, feature sets, classification techniques and evaluation strategies. We have done a comprehensive study of the existing research and gave our own inference on the topic. In this paper we discuss where the Keystroke Dynamics research currently stands and what scope it has in the future as a biometric application.

Keywords - Keystroke Dynamics, Behavioral Biometrics, User Authentication, Identification, Computer Security.

32. PaperID 310316131: A Novel Supervised Approach to Detection of Shilling Attack in Collaborative Filtering Based Recommendation System (pp. 208-211)

*Krupa Patel, Department of Information Technology, CSPIT, CHARUSAT, Anand, India.
Amit Thakkar, Associate Professor, Department of Information Technology, CSPIT, CHARUSAT, Anand, India.
Chandni Shah, Assistant Professor, Department of Information Technology, CSPIT, CHARUSAT, Anand, India.
Kamlesh Makvana, Assistant Professor, Department of Information Technology, CSPIT, CHARUSAT, Anand, India.*

Abstract — Collaborative filtering is widely used recommendation algorithm to generate variety of recommendation for target users. With increasing popularity of collaborative filtering recommendation, number of users started to insert fake shilling profiles into the system. Due to shilling attack or profile injection attack, accuracy of collaborative filtering recommendation will reduce. This paper attempts to proposed method to detection of shilling attack in collaborative filtering recommendation system using supervised approach. Our proposed method use statistical parameters RDMA, DigSim and LengthVar to identify shilling attack profiles from genuine profile. This parameters are used to train the model for detection of attacker profiles. Then our proposed method will identify genuine profile those are classified as attacker profiles.

Keywords — Recommendation System, Collaborative Filtering, Shilling Attack, Profile Injection Attack, Supervised Learning, Statistical parameters.

33. PaperID 310316140: Privacy Preserving Data Classification using a New Heterogeneous Data Distortion (pp. 212-217)

J. Hyma (†), PVGD Prasad Reddy (††), and A. Damodaram (†††)

† Department of CSE, GIT, GITAM University, Visakhapatnam, INDIA

†† † Department of CS&SE, AU College of Engineering, Andhra University, Visakhapatnam, INDIA

††† Department of CSE, Sri Venkateswara University, Tirupathy, INDIA

Abstract - The new digital technology facilitates us to collect huge amount of data every day. Due to this tremendous growth in size and complexity, two important factors have got the increased attention of all the technology users. One is the complex data analysis that could be done using various data mining methods. The second is privacy concern of the individual towards their data. Privacy Preserving Data Mining (PPDM) is one such process that pays an equal attention towards these two factors. Though there are various techniques in PPDM process, there is no such existing technique that exerts the equal amount of importance on all the roles involved in communication. Our proposed model not only considers the various roles like data owners, data collectors and data users, but also applies the required set of heterogeneous constraints to obtain better privacy protection and better data usability. Heterogeneous constraints used in this work are proposed basing upon the owners willingness to publish the data and existing correlations and privacy analysis carried out by the anonymization framework of the data collector layer.

Keywords: Privacy preserving data mining (PPDM), Heterogeneous constraints, Privacy preserving data classification.

34. PaperID 29021627: Evaluating the Effects of Network Size and Nodes Mobility Speed on the Performance of TCP over Mobile Ad-Hoc Networks (pp. 218-227)

O. G. Aju, O. Oriola

Department of Computer Science, Faculty of Science, Adekunle Ajasin University, Akungba-Akoko, Ondo State, Nigeria

Abstract — The purpose for the design of Transmission Control Protocol (TCP) was to provide reliable end-to-end delivery of data over unsecured networks. Although, it is designed to be deployed in the traditional wired networks but recently, there has been an increase in its deployment over the wireless networks such as Mobile Ad-Hoc Networks (MANETs). This paper investigates the performance of various TCP variants in specified network scenarios in Mobile Ad hoc Networks (MANETs) using Reno, New Reno and SACK as case study under the Dynamics Source Routing (DSR) Protocol by observing the effects of some network designs on the performance of TCP variants in MANETs using throughput, delay and retransmission attempts as performance metrics. Application traffics were submitted to MANETs while the network size (number of nodes) and the nodes mobility speed were varied to create network models and the resulting throughput, end-to-end delay and retransmission attempts were observed to determine how the network size and the nodes mobility speed affects the performance of the TCP variants.

Index Terms— Mobile Ad hoc Network, Transmission Control Protocol, Selective Acknowledgements, File Transfer Protocol, Hypertext Transfer Protocol, Voice over Internet Protocol.

35. PaperID 29021672: Adaptive Fuzzy Control to Design and Implementation of Traffic Simulation System (pp. 228-246)

Laheeb Mohammed Ibrahim, Software Engineering, Mosul University, Collage of Computer Sc. & Math., Mosul, Iraq

Mohammed A. Aldabbagh, Software Engineering, Mosul University, Collage of Computer Sc. & Math., Mosul, Iraq

Abstract — In this paper a Fuzzy Adaptive Traffic Signal System (FATSS) was designed and implemented to improve optimization and compare fix time traffic light controller. FATSS allows the user to select input parameters and tune rule base to improve optimization and compare fix time traffic light controller. FATSS reducing the average waiting time for vehicles between 2% to 20%, and that indicate the adaptive traffic light controller based on fuzzy logic outperform is better when is compare with other fixed controller FATSS was built using C# language in Microsoft Visual studio 2010 development environment. The simulation is implemented by Simulation for Urban Mobility (SUMO).

36. PaperID 29021695: Secure Cooperative Intrusion Detection System for Mobile Ad-Hoc Network (pp. 247-250)

*Himanshu Kumar, Department of Information Technology, SRM University Kattankulathur Chennai
J. Godwin Ponsam, Asst. Professor, Department of Information Technology, SRM University Kattankulathur, Chennai*

Abstract - The Mobile Ad-Hoc Network does not have any fixed infrastructure so they rely on their neighbors to relay data packets over a network. Intrusion detection system in mobile ad-hoc network can be carried out in a distribution scenario due to absence of fixed infrastructure. This nature of MANET attracts the malicious users. Intrusion Detection System are the techniques to detect the malicious node. The objective of this project is to propose an Energy efficient system based on a cooperative IDS scheme to deal with intrusions in clustered mobile ad-hoc networks. We are analyzing the Energy Consumption of MANET by using present Protocols in terms of Packet dropping detection ratio, Mobility stability and Transmission Power Control etc.

Keywords: Ad-hoc Network, IDS, Energy Consumption, MANET, Wireless Network

37. PaperID 310316119: Ensuring Interoperability among Heterogeneous Devices through IoT Middleware (pp. 251-255)

*Muhammad Ahsan, M. Ramzan Talib, M. Umer Sarwar, M. Irfan Khan, M. Bilal Sarwar
Department of Computer Science, Government College University Faisalabad, Pakistan*

Abstract — Internet of Things provides truly ubiquitous and smart environment. The multilayer distributed architecture with a variety of different components together with end devices, applications and the association with its framework poses challenge. Internet of Things middleware actions as a joining link between the heterogeneous areas that communicate across heterogeneous edges. In this work, we study the interoperability issue between heterogeneous devices. We presented guidelines to handle the interoperability issue in Internet of Things. Furthermore, we have proposed architectural framework for Home Area Network.

Keywords - Interoperability, Internet of things, Middleware, Heterogeneous devices

38. PaperID 31031699: SQL Injection Attack Detection & Prevention over Cloud Services (pp. 256-261)

*Niharika Singh, Ajay Jangra, Upasana Lakhina, Rajat Sharma
Department of Computer Science and Engineering, University Institute of Engineering and Technology, Kurukshetra University, Kurukshetra, India*

Abstract — Web servers which provide customer services are usually connected to highly sensitive information contained backend databases. The incrementing bar of deploying such web applications initiated in ranging the corresponding bar of number of attacks that target such applications. SQL Injection Attacks come about when data provided by external user are directly included in SQL query but is not properly validated. The paper proposes a novel detection & a prevention mechanism of SQL Injection Attacks using three-tier system. As the methodology is concerned over static, dynamic & runtime detection and prevention mechanism which also filters out the malicious queries and inspires the system to be well prepared for the secure working environment, regardless of being concerned over the database server only. The cloud proposes the services like SaaS, IaaS, PaaS, DaaS, EaaS. As previous solutions are achieved for the database queries for DaaS service only, but this paper enhances the scope of other services as well. It adapts to maintain security of the whole system even when it is for any of the cloud platforms. The solution includes detection & filtration that reduces attacks to 80% in comparison to other algorithms.

Keywords — Cloud computing; Cloud Security; Architecture, design; Cloud services; Deployment models; SQL Injections;

39. PaperID 310316125: Survey on Issues in Wireless Sensor Networks: Attacks and Countermeasures (pp. 262-269)

*Rangstone Paul Kurbah, Bobby Sharma
Assam Don Bosco Unniversity, Guwahati, Assa*

Abstract — Wireless Sensor Networks have become popular day by day. They find their applications in numerous areas. These networks, however, have some constraints like the physical size (that they must be compact), energy (that minimum energy must suffice them for long hours), memory space (that they should effectively work with just minimum memory space installed on them), and above all that their construction cost must be minimum. Due to these constraints they face some security issues. Securing the data that flows through these networks must be of paramount importance and the security issues that are faced by these networks must be addressed in order to enhance their reliability and usage. This paper focuses on the security aspects of Wireless Sensor Networks. It presents the general characteristics of Wireless Sensor Networks, their constraints, their security goals, the threat models, the different types of attack on WSNs and their defensive measures.

Keywords: Attacks, Defensive Measures, Nodes, Security, Wireless Sensor Network (WSN).

40. PaperID 29021641: Syntactical Knowledge and Sanskrit Memamsa Principle Based Hybrid Approach for Text Summarization (pp. 270-275)

*D. Y. Sakhare (1, 2), Raj Kumar(3)
(1) Bharativedyapeeth Deemed University's College of Engineering, Pune, MS, India
(2) MIT Academy of Engineering, Alandi, Pune, MS, India
(3) DIAT, Khadakwasala Pune, MS, India*

Abstract - The proposed approach works towards integrating syntactic knowledge and sentence fusion for abstractive multi-document summarization system. A fuzzy logic system, based on the “Paninian” Parts of Speech Tagging, is used to extract the syntactical knowledge-based informative words from English the documents. The sentences containing the informative words are selected for the further processing of abstractive summary generation. The sentence formation for the abstractive summarization is done using a neural network with features based on the Memamsa principles of the Sanskrit language. The features, such as “Upakram-Upsanhar,” “Abhyas,” “Apurvata,” “Phalam,” “Sthan,” “Prakaran” and “Samakhya” are used to form meaningful sentences. These features and the target summary of each document are given as input to train the neural network. The neural network trains the system based on the target summary of a set of documents with the same information to generate an abstractive summary for a new cluster of documents. The system performance is measured on a real data set and the DUC 2002 data set using ROUGE-1 and ROUGE-2 scores and the F-measure. The proposed Fuzzy- NN approach performs better than the existing techniques.

Keywords: Text summarization, Informative Word, Sentence Formation, Memamsa principles, Fuzzy NN, ROUGE

41. PaperID 31031602: The Smartcane for Blind People an Electronically Smart Stick to Aid Mobility (pp. 276-285)

*M. Asad Hussain, M. Ghazanfar Ullah, Atif Fareed, Basit Sohail
Department of Electrical Engineering, Usman Institute of Technology – Pakistan*

Abstract — This paper is focused about the development of a — Micro-controller based Smart White Cane || A.K.A. —The Smartcane || and its comparison, based on performance and usability, with other existing models. Our main contribution is to enhance the capabilities of existing models of micro-controller based white stick for blind persons, due to their practical limitations. The developed project serves the best solution to overcome the difficulties of blind people, so that they can easily mobilize themselves, be a more successful part of society. The developed

project facilitates blind persons in a manner that they can handle any obstacle, wet material, uneven surface, etc. Our main objective was to reduce the size of the presented model by integrating the circuits and making it a compact and portable stick for users. Also, we emphasize on the range of the modules and sensors to increase the efficiency and usability of the prototype model. The system accompanied a portable unit that can easily be carried and operated by a visually impaired user. It could easily be incorporated into a walking cane. The salient features of the developed prototype are ultrasonic sensor for obstacle detection, water probe for mud and water detection, I.R. for ditch detection, G.P.S, G.S.M. module, signal-to-speech module, speaker or headset, and portability (size and power). The experimental results shows that the developed prototype is much more efficient and usable in varying situations for a blind person as compared to the ordinary white sticks while affordable and cost effective at the same time.

Keywords – Blind, Mobility Aid, Smartcane, Microcontroller, GPS, GSM, Ultrasonic sensor, IR sensor.

42. PaperID 31031612: E-Commerce Framework Based on Evaluation of Data Mining and Cloud Computing (pp. 286-295)

*Mohd Muntjir, College of Computers and Information Technology, Taif University, Taif, Saudi Arabia
Ahmad Tasnim Siddiqui, College of Computers and Information Technology, Taif University, Taif, Saudi Arabia*

Abstract - This paper is a description about the application of e-commerce and data mining with cloud Computing. It emphasizes how data mining is used for e-commerce in combination of cloud computing systems. Data Mining is a process of separating possibly useful information from available raw data. It's also describing that How SaaS is very useful in cloud computing. The combination of data mining techniques into normal day-to-day actions has become common part. Businesses and advertising have become more active through the use of data mining functionalities to deduct the overall costs. Data mining operations can develop much more demographic information respecting customers that was basically not known or hidden in the desired data. It has basically seen enhancements in data mining techniques proposed to such activities as identifying criminal activities, fraud detection, suspects, and indication of potential terrorists. On the whole, data mining systems that have been designed and developed to data for grids, clusters, and distributed clusters have considered that the processors are the limited resource, and hence distributed. When processors become accessible, the data is transferred to the processors.

Keywords: Data Mining, e-commerce, cloud computing systems, data mining and cloud computing, (SaaS) Software-as-a-Service.

43. PaperID 31031618: Radial Basis Polynomial Kernel (RBPk): A Generalized Kernel for Support Vector Machine (pp. 296-315)

*Ms. Hetal Bhavsar (1), Dr. Amit Ganatra (2)
(1) Assistant Professor, Department of Computer Science & Engineering, The M. S. University of Baroda, Vadodara, India
(2) H. O. D., Computer Engineering Department, Charotar University of Science & Technology, Changa, Dist. Anand, India.*

Abstract - Support Vector Machine (SVM) is a novel machine learning method, based on the statistical learning theory and VC (VapnikChervonenkis) dimension concept. It has been successfully applied to numerous classification and pattern recognition problems. Generally, SVM uses the kernel functions, when data is non-linearly separable. The kernel functions map the data from input space to higher dimensional feature space so the data becomes linearly separable. In this, deciding the appropriate kernel function for a given application is the crucial issue. This research proposes a new kernel function named — Radial Basis Polynomial Kernel (RBPk) || which combines the characteristics of the two kernel functions: the Radial Basis Function (RBF) kernel and the Polynomial kernel and proves to be better kernel function in comparison of the two when applied individually. The paper proves and makes sure that RBPk confirms the characteristics of a kernel. It also evaluates the performance of the RBPk using Sequential Minimal Optimization (SMO), one of the well known implementation of SVM, against the existing kernels. The simulation uses various classification validation methods viz. holdout, training vs. training, cross-validation and random sampling methods with different datasets from distinct domains to prove the usefulness of RBPk. Finally, it

concludes that the use of RBPK results into better predictability and generalization capability of SVM and RBPK can become an alternative generalized kernel.

Keywords: Support vector machine; kernel function; sequential minimal optimization; feature space; polynomial kernel; and Radial Basis function

44. PaperID 31031620: MOiD (Multiple Objects incremental DBSCAN) – A paradigm shift in incremental DBSCAN (pp. 316-346)

Neha Soni (1), Dr. Amit Ganatra (2)

(1) Computer Engineering Dept., SVIT, Vasad, Gujarat, India

(2) Faculty of Tech. & Engg., CHARUSAT, Changa, Gujarat, India

Abstract - Mining an unprecedented increasing volume of data is a herculean task. Many mining techniques are available and being proposed every day. Clustering is one of those techniques used to group unlabeled data. Among prevailing proposed methods of clustering, DBSCAN is a density based clustering method widely used for spatial data. The major problems of DBSCAN algorithm are, its time complexity, handling of varied density datasets, parameter settings etc. Incremental version of DBSCAN has also been proposed to work in dynamic environment but the size of increment is restricted to one data object at a time. This paper presents a new flavour of incremental DBSCAN which works for multiple data objects at a time, named MOiD (Multiple Objects incremental DBSCAN). MOiD has been experimented on thirteen publicly available two dimensional and multi-dimensional datasets. The results show that MOiD performs significantly well in terms of clustering speed with a minor variation in accuracy.

Keywords - Incremental Clustering, DBSCAN, Density based clustering, region query, clustering

45. PaperID 31031629: Wavelet based OFDM with ICI Self Cancellation for Underwater Acoustic Communications (pp. 347-352)

Naresh Kumar, Member, IEEE and B. S. Sohi, Sr. Member, IEEE

University Institute of Engineering & Technology (UIET), Panjab University Chandigarh, India

Abstract — There are many research challenges in underwater acoustic communication environment such as large delay spread, ocean waves, motion of transmitter/receiver, Doppler spread etc. OFDM has potential to combat with many such problems, but it is also deteriorated by Inter Carrier Interference and high peak to average power ratio. Conventional OFDM is spectral inefficiency as it uses cyclic prefixing which consumes approximately 20% of available bandwidth. ICI self-cancellation technique performs better for ICI problems. As it transmits redundant data on adjacent subcarriers which makes some subcarriers idle, hence, ICI is reduced at the cost of bandwidth. In this paper, a Wavelet based OFDM with ICI cancellations is proposed to counter the problem of ICI. Use of Wavelets reduces the need for cyclic prefixing thereby making it more spectral efficient and wavelets also help in maintaining orthogonality between subcarriers which further improves its ICI performance. Simulation results show that proposed technique performs better in terms of bit error rate (BER) as compared to conventional OFDM.

Index Terms — OFDM, Wavelets, BER, Self-Cancellations, ICI

46. PaperID 31031632: A Method for Mining Social Media to Discovering Influential Users (pp. 353-365)

Hosniyeh S. Arian (1), Omid R. B. Speily (2)

(1) Department of Computer Engineering and Information Technology, Islamic Azad University, Qazvin, Iran.

(2) Department of Computer Engineering and Information Technology, Urmia University of Technology, Urmia, Iran

Abstract - Influential users who diffuse information and their followers have interest to this information finally they can maximize diffusion in social networks. Influential users have different influence in diversity domain specificity

for instance user may have strong influence in a special topic and another topics have weak influence. So a proposed method presented for identifying influential users based on domain specificity in this paper. This method identified influential users based on domain specificity that features of user's profile and user's actions (e.g. retweet) that influence on diffusion determined by "multiple regression" and user's contents categorized based on keywords by "TF-IDF" and finally influential users identified by "Tree Regression" based on domain specificity in this paper. The detail of this method discussed the following of paper. In order to evaluate the proposed method on Twitter offer application program interface. 420 users selected randomly, they follow their friends, join to different groups, and generated diversity tweets on Twitter. The main feature, which distinguishes this method from the previously reported methods, is in two key respective. First previous studies have quantified influence in terms of network metrics for instance number of retweet or page rank, our proposed method measured influence in terms of the size Tree Regression. Second the focuses of previous studies were based on the structural of diffusion and feature of content but Influential users have different influence in diversity domain specificity so in our proposed method focused on this feature. Results showed that accuracy of proposed method is 0.69.

Keywords: Social networks, Categorized, Influence, Content, Diffusion, Domain specificity.

47. PaperID 31031642: StudentPIMS: A Smartphone-Based Personal Information Management System for Students (pp. 366-380)

Irfan Ullah (a,b,), Shah Khusro (b), Habib Un Nabi (a), Rafi Ullah (a)
(a) Department of Computer Science, Shaheed Benazir Bhutto University, Sheringal, 18050, Pakistan
(b) Department of Computer Science, University of Peshawar, Peshawar, 25120, Pakistan*

Abstract - Curricular and co-curricular activities are among the major responsibilities that require proper attention from the students in order to achieve different goals and objectives regarding their bright future. Because of the mismanagement of keeping personal information about these activities, most of students are unable to remember these tasks while they are busy in their studies and therefore, fail to perform these activities at the right time. To handle this issue, they adopt several means including SMS drafts, reminders, sticky notes, notebooks, dairies, and laptops etc., which are limited and unable to fully support students because of several problems including their storage, search, and retrieval. With the availability and wide-spread adaptation of Android and Smartphones, researchers and developers started thinking of new and innovative ways of managing personal information of people especially students. Today, several apps are available on Google Play for managing personal information of students. However, the existing solutions have limitations including bulky user interfaces especially when the stored information exceeds a certain limit, usability, privacy, and requiring access to Internet for accessing certain services, which becomes a barrier to students especially to those living in rural areas of developing countries where access to Internet is among the major issues. Keeping in view these limitations, we have designed and developed StudentPIMS - a simple and usable Android app that allows students to easily manage personal information about these activities without suffering from cognitive overload caused by existing solutions. We have compared our solution with the existing solutions using some evaluation metrics as well as conducted a survey research among users of the app. Results show that StudentPIMS outperforms the available solutions especially in terms of usability, privacy, and low resource consumption.

48. PaperID 31031652: Contribution to a proportional sharing out of IaaS resources for service adaptation with users services according its profile in cloud computing (An equity based approach) (pp. 381-395)

*KANGA Koffi, Ecole Doctorale Polytechnique de l'Institut Nationale Polytechnique Félix Houphouët Boigny (EDP/INPHB), Côte D'ivoire UMRI 78 Laboratoire de recherche en informatique et télécommunication
GOORE Bi Tra, Institut Nationale Polytechnique Félix Houphouët Boigny (INPHB), Côte D'ivoire, Laboratoire de Mathématiques et des Nouvelles Technologies de l'Information
BABRI Michel, Institut Nationale Polytechnique Félix Houphouët Boigny (EDP/INPHB), Côte D'ivoire, Laboratoire de Mathématiques et des Nouvelles Technologies de l'Information*

Abstract - Cloud computing ensures the allowance of resources consumption to the user, by paying for it as he will do for other basic services as water and electricity. In this article we propose an IaaS resource adaptation technique (space

capacity) necessary for the SaaS and PaaS in order to improve their functioning in terms of storage capacity by taking into account users' profile. In that way, a proportionality coefficient has been defined, and used for this adjustment and also by taking into account previous IaaS space occupations proportion for each service of cloud. Our contribution is based on the setting up of an allocation technique supported by an algorithm allowing its achievement. The outcome results of the implementation of the algorithm show that our method allows a propositional sharing out of the resources. Therefore the IaaS space should be adapted to the users' service.

Keywords: Cloud computing, Users profile, resources allocation, IaaS resources adaptation.

49. PaperID 31031654: Enhanced Data Security in Network Environment (pp. 396-405)

Ram Krishna Akuli, Dr. J. Durga Prasad Rao, Dr. Satyendra Kurariya

(1) Scholar CVRU, Bilaspur

(2) Additional Director & HOD (Computer Science Department), Shri Shankaracharya Mahavidyalaya, Junwani, Bhilai

(3) Head, Computer Science Department, Mata Gujari College, Jabalpur

Abstract - This study is based on the development of a new secure protocol for remote calls. The secure protocol design specification and descriptions are analysed comparing them with the existing protocols. The protocol is designed in a simple way with in built security features. Cryptographic modules can be exchanged due to the flexibility of the new approach depending on various issues and security matters. The developed protocol in this study is platform independent. The security levels of the new secure protocol are properly analysed with desired results. Comparisons with other existing technologies like CORBA or the RMI were also addressed. The results show that creation of a secure network protocol universally acceptable. Although all the bugs and security issues were not addressed as they keep evolving on a daily basis.

Keywords: - Cryptographic Protocol, Secure Remote Protocol, Network Security

50. PaperID 31031661: Security Concerns with Open Research Issues of Present Computer Network (pp. 406-432)

Geetanjali Rathee, Hemraj Saini

Department of Computer Science and Engineering, Jaypee University of Information Technology, Wknaghat, Solan-173234, Himachal Pradesh, India

Abstract - Present networks are the mainstay of modern communication. The existence of networks is enriching our society in countless different ways. Now days, wireless mesh network is considered as an auspicious technology for posing self-healing, organizing and configurable capabilities but one of the foremost challenge in the enterprise of these networks is their susceptibility to security assaults (eavesdropping, network layer attacks and denial of service). In order to overcome against these assaults, several security anxieties are proposed but authentication is taken as an important parameter to provide a secure communication. In this chapter, a review is discussed from origin to the current networking technology i.e. WMN. In addition to this, WMN security is concerned with recent applications such as smart grids, intelligent transportation system, multimedia systems etc. further a clear overview of security with respect to each layer is elucidated and finally the chapter is ruined by outlining the future work which is the next step of this research

51. PaperID 31031662: Performance-aware Cloaking Algorithm for Optimizing Anonymous Location-based Services (pp. 433-439)

Dileep Kumar, Department of Information Media, The University of Suwon, Hwaseong-si South Korea

Abstract - The prevailing infrastructure of ubiquitous computing paradigm on the one hand making significant development for integrating technology in the daily life but on the other hand raising concerns for privacy and

confidentiality. As Location based services (LBS) equip users to query information specific to a location with respect to temporal and spatial factors thus LBS in general while Location Anonymizer, core component of privacy preservation models, in particular put under extreme criticism when it comes to location privacy, user confidentiality and quality of service. For example, a mobile or stationary user asking about his/her nearest hospital, hotel or picnic resort has to compromise their exact location information. Here in this paper we are addressing the significance of our proposed index optimized cloaking algorithm for Location Anonymizer with respect to performance, quality and accuracy which can be smoothly integrated into existing location anonymity model for privacy preservation. The main idea is to deploy R-tree based indexing scheme for Location Anonymizer to make best use of available computing resources. In accordance with the proposed approach, next step is to develop an index optimized cloaking algorithm which can cloak spatial region effectively and efficiently on behalf of R-tree based indexing scheme. Finally we will quantify the benefits of our approach using sampled results through experiments that the proposed cloaking algorithm is scalable, efficient and robust to support spatio-temporal queries for location privacy.

52. PaperID 31031671: Encrypting Grayscale Images using S8 S-Boxes Chosen by Logistic Map (pp. 440-444)

Tariq Shah, Ayesha Qureshi

Department of Mathematics, Quaid-i-Azam University, Islamabad, 44000, Pakistan

Abstract - In the present manuscript, we will design an encryption algorithm for grayscale images that is based on S8 S-boxes transformations constructed by the action of symmetric group S8 on AES S-box. Each pixel value of the plain image is transformed GF (2^8) into with a dissimilar S8 S-box chosen by using the logistic map. In this way, there are 40,320 possible choices to transform a single pixel of the plain image. By applying the generalized majority logic criterion, we will establish that the encryption characteristics of this approach are superior to the encoding performed by AES S-box or a single S8 S-box.

Keywords: AES S-box, S-boxes, logistic map, generalized majority logic criterion.

53. PaperID 31031677: A Dynamic Media Access Control Protocol for Controlling Congestion In Wireless Sensor Network By Using Fuzzy Logic System And Learning Automata (pp. 445-460)

Foroogh Karimi, Mansour Esmailpour

Department of Computer Engineering, Hamedan Branch, Islamic Azad University, Hamedan, Iran

Abstract - One of the existing layers in the reference model whose designing is of particular complication is the control layer of access to MAC media, it's proper designing causes to reduce interference and consequently to reduce energy consuming and to increase the network efficiency. In the recommended method, our focus is on the networks being multi-channel in order to distribute the network current through the different channels. In the first step of the research, we have used a layering structure for a better management of the network so that we could prevent congestion via the network management. This management is performed through using Fuzzy logic system logic system. The output of our Fuzzy logic system is the election of the best and most appropriate choice in order to continue route finding. But if a congestion of one incident takes place, we possess learning automata for assigning the channel searchingly for balancing the channel current. Using the resemblance maker of NS2, the results of the resemblance-making maintain that the recommended method has improved more greatly than the two basic protocols and could achieve the quality parameters of route finding services.

Keyword: Wireless sensor networks, Congestion control, Multichannel, Fuzzy logic system, Learning Automata

54. PaperID 310316103: Presenting a Model to Meet the Pathology Component in the Implementation of Beyond Component Processes in Distributed Integrated Information Systems (pp. 461-470)

Masoud Rafighi, Yaghsoub Farjami

Department of Computer Engineering and Information Technology, University of Qom, Qom, Iran

Abstract - making all the applications in an enterprise work in an integrated manner, so as to provide unified and consistent data and functionality, is a difficult task because it involves integrating applications of various kinds, such as custom-built applications (C++/C#, Java/J2EE), packaged applications (CRM or ERP applications), and legacy applications (mainframe CICS or IMS). Furthermore, these applications may be dispersed geographically and run on various platforms. In addition, there may be a need for integrating applications that are outside the enterprise. According the problems of adding application to organization and keep integration between them, in this paper, we studied the ways of integration between systems of organization. Then consider the Problems of models and emphasize on crucial need to create an ideal model for optimal architecture which meets the needs of the organization for flexibility, extensibility and integration of systems. Finally proposed a model which in addition doing comprehensive processes between the components easily in distributed systems, it does not have the problems of previous models. Since components are vulnerable in sending beyond component processes, so in this article we decided to introduce a model of pathology components to resolve the implementation of beyond component processes.

Keywords: ESB, Data-centric architecture, architecture Component-based, Plug in architecture, distributed systems.

55. PaperID 310316112: A Novel Energy Efficient Connected Target Coverage Heuristic in WSN (pp. 471-479)

Sunita Gupta, Ph.D. Scholar, Suresh Gyan Vihar University, Jaipur

Dr. K. C. Roy, Professor, Kautilya Institute of Technology & Engineering, Jaipur

Dr. Dinesh Goyal, Professor, Suresh Gyan Vihar University, Jaipur

Sakar Gupta, Associate Professor, Kautilya Institute of Technology & Engineering, Jaipur

Abstract - Wireless Sensors Networks (WSNs) are able to work in insensitive environments where real observations by human being are dangerous, incompetent and sometimes not feasible. A most significant characteristic of a WSN application is lifetime. Wireless sensor network can be used till they can sense and communicate the sensed data to base station. Sensing as well as communication, both are important functions and they use energy. Energy management and scheduling of sensors can effectively help in rising the networks lifetime. Energy efficiency in a region monitored by a sensor network is achieved by dividing the sensors into cover sets. Every cover set is able to monitor the targets for a definite time period. At a time only single cover set is in active state and rest others are in low power sleep state. Thus energy is preserved and lifetime of Wireless Sensor Network is increased. Creating the greatest number of such set covers is proved to be an NPC problem. An energy minimization heuristic called Q-Coverage P-Connectivity Maximum Connected Set Cover (QC-PC-MCSC) is proposed. Functioning of Sensor nodes is scheduled in such a manner that they are having Q-Coverage and P-Connectivity constraint and thus they improves the working duration of Wireless Sensor Network. A comparative study of performance of QC-PC-MCSC and existing heuristic is also done over Energy Latency Density Design Space for Wireless Sensor Network.

Keywords:- Wireless Sensor Network, Connected Target Coverage, Network Lifetime, Cover Set, Coverage, Connectivity, Q-Coverage, P-Connectivity.

56. PaperID 310316115: A Novel Hybrid Encryption Scheme to Ensure Hadoop Based Cloud Data Security (pp. 480-484)

Danish Shehzad (1), Zakir Khan (2), Hasan Dağ (3), Zeki Bozkuş (4)

(1, 4) Department of Computer Engineering, (3) Department of Management Information Systems, Kadir Has University, Istanbul, Turkey

(2) Department of Information Technology, Hazara University, Mansehra, Pakistan

Abstract - Cloud computing and big data have provided a solution for storing and processing large amount of complex data. Despite the fact that they are quite useful, the threat to data security in cloud has become a matter of great concern. The security weakness in Hadoop, which is an open source framework for big data and cloud computing, has setback its deployment in many operational areas. Different symmetric, asymmetric, and hybrid encryption schemes have been applied on Hadoop for achieving suitable level of data security. In this paper a novel hybrid encryption scheme, which combines symmetric key algorithm using images as secret keys and asymmetric data key encryption using RSA, is proposed. The suggested scheme reduced the overhead of the secret key computation cycles as compared to the other

existing encryption schemes. Thus, it is safe to claim that the proposed scheme retains adequate security level and makes data encryption more efficient.

Keywords: Hadoop, Hadoop distributed file systems (HDFS), Matlab, Data encryption scheme (DES), RSA.

57. PaperID 310316118: Enhancing Users' Satisfaction Using an Agent-Based Negotiation System (pp. 485-496)

Omid R. B. Speily (1), Yosra Bahrani (2), Negin Razavi Rajayi (2)

(1) Department of Information Technology & Computer Engineering, Urmia University of Technology, Urmia, Iran

(2) Department of Information Technology & Computer Engineering, AmirKabir University of Technology, Tehran, Iran

Abstract - The increasing importance of operating automated systems arises with emerging competitive e-commerce environment. Nowadays, operating automated systems used in conducting all business transactions are enhanced substantially to achieve beneficial trade and decrease frequent messaging overhead of transactions. In spite of the highly competitive electronic marketplace, it is necessary to design a system which automates tasks including group negotiation and, payment and delivery. In this paper, we apply the purchasing groups to enhance the bargaining power of customers still satisfying all users' needs and preferences. We propose a flexible system called UUT-Trade to purchase laptop computers. This system uses a novel negotiation algorithm which diminishes all prices offered by potential sellers as much as possible, and then users will have the chance to choose between potential sellers by performing a weighted voting. Unlike similar systems which also exploit group purchasing, this system suggests no scarification of buyers' needs.

Keywords: Negotiation, Automation, Scarification, UUT-Trade, AHP tree.

58. PaperID 310316123: Parallelizing K-Way Merging (pp. 497-503)

H M Bahig (1, 2) and Ahmed Y Khedr (1, 3)

(1) College of Computer Science and Engineering, Hail University, Hail, KSA

(2) Department of Mathematics, Faculty of Science, Ain Shams University, Cairo, Egypt

(3) Systems and Computer Department, Faculty of Engineering, Al-Azhar University, Cairo, Egypt

Abstract — The k-way merging problem is to find a new sorted array as an output from k sorted arrays as an input. In this paper, we consider the elements of the k sorted arrays are data record, where the value of the key for each record is a serial number. The problem is used to design efficient external sorting algorithm. We proposed two optimal parallel algorithms for k merging. The first one is based on merging k sorted arrays of n records in a new sorted array of length n. The second one is based on merging k sorted arrays of n records in a new sorted array of length $n+o(n)$ which is called padded merging. The running time for each algorithm is $O(\log n)$ and $O(1)$ under EREW and CRCW PRAM respectively.

Keywords - merging; k-merging; padded merging; PRAM; optimal algorithm; parallel algorithm.

59. PaperID 310316130: Extended Smart Metering Display for Improved Energy Economy (pp. 504-512)

Nisar Ahmed (1), Muzafar Khan (2), Muhammad Tahir (3), Shahid Yousaf (1)

(1) School of Engineering, Blekinge Institute of Technology, Karlskrona, Sweden

(2) College of Computer and Information Sciences (Muzahmiyah Branch), King Saud University, Riyadh, Saudi Arabia

(3) Faculty of Computing and Information Technology, University of Jeddah, Jeddah, Saudi Arabia

Abstract - Human dependency on technology is increasing day by day and environmental conditions are getting worse as a result. Energy consumption is increasing while the traditionally available energy sources like oil and gases are depleting. One of the major consumers is the domestic consumer, who plays the least part in energy management. One

way to increase efficiency in energy management is, therefore, to pass part of it to the domestic consumer, what is known as self-management. For the consumers to do self-management, they require the relevant information pertaining to their consumption patterns. Smart heat meters are already being used to provide this information. However, they are still being under-utilized in terms of their capability. In this research work an Extended Smart Metering Display (ESMD) is proposed; it is based on the interviews conducted with the representatives of smart heat meter manufacturers, District Heating (DH) providers and domestic consumers of DH in the Blekinge county of Sweden. The proposed ESMD was evaluated by the member companies of Swedish District Heating Association and domestic consumers in the workshop conducted for this purpose. The proposed ESMD may help the domestic consumers in monitoring their energy consumption on real-time basis, and improving their energy consumption behavior. It is also suggested that how it can be made more financially viable for the energy consumers and providers during the peak hours, if the proposed system is used.

Keywords: consumer behavior measurement, district heating, energy economy, metering display, smart heat meter

60. PaperID 310316132: Classifying Facial Expressions using DWT, Moments and Genetic Algorithm (pp. 513-522)

M. Mahadevi (1), Dr. C. P. Sumathi (2)

(1) Research Scholar (M. S. University) & Asst. Professor, SDNB Vaishnav College for women, Chennai, Tamilnadu, India

(2) Associate Professor & Head, Department of computer science, SDNB Vaishnav College for women, Chennai, Tamilnadu, India

Abstract - Facial expressions are the actions of the thoughts that arise in a mind. Such expressions are categorized as simple basic and complex expressions which are a mixture of two or more expressions. This research focuses on identifying the basic expressions and classifying them based on Naïve Bayes classifier. The database considered for the research is Japanese Female Facial Expression (JAFPE) consisting seven expressions happy, sad, disgust, fear, angry, neutral and surprise. The image is pre-processed using Discrete Wavelet Transform (DWT) and created a feature set containing spatial statistical features of the facial parts and moments of the DWT image. The features were selected using genetic algorithm and classified the database using Naïve Bayes classification to acquire an overall accuracy rate of 92.5%.

Keywords: Spatial Statistical features, DWT, Genetic algorithm, Naïve Bayes

61. PaperID 310316141: A Modern Approach to Integrate Database Queries for Searching E-Commerce Product (pp. 523-531)

Ahmad Tasnim Siddiqui, College of Computers and Information Technology, Taif University, Taif, Saudi Arabia

Mohd. Muntjir, College of Computers and Information Technology, Taif University, Taif, Saudi Arabia

Abstract - E-commerce refers to the utilization of electronic data transmission for enhancing business processes and implementing business strategies. Explicit components of e-commerce include providing after-sales services, promoting services/products to services, processing payment, engaging in transaction processes, identifying customer's needs, processing payment and creating services/products. In recent times, the use of e-commerce has become too common among the people. However, the growing demand of e-commerce sites have made essential for the databases to support direct querying of the Web page. This re-search aims to explore and evaluate the integration of database queries and their uses in searching of electronic commerce products. It has been analyzed that e-commerce is one of the most outstanding trends, which have been emerged in the commerce world, for the last decades. Therefore, this study was undertaken to ex-amine the benefits of integrating database queries with e-commerce product searches. The findings of this study suggested that database queries are extremely valuable for e-commerce sites as they make product searches simpler and accurate. In this context, the approach of integrating database queries is found to be the most suitable and satisfactory, as it simplifies the searching of e-commerce products.

Keywords: E-commerce product search, e-commerce, query optimization, business processes, Query integration

62. PaperID 310316144: Evaluation of Machine Vision for Robot Navigator: Mini Review (pp. 532-540)

Arefe Esalat Nejad

Young Researchers and Elite Club, Baft Branch, Islamic Azad University, Baft, Iran

Abstract - Machine vision (MV) is the technology and methods used to provide imaging-based automatic inspection and analysis for such applications as automatic inspection, process control, and robot guidance in industry. This paper presents some of the underlying concepts and principles that were key to the design of our research robots. Vision is an ideal sensor modality for intelligent robots. It provides rich information on the environment as required for recognizing objects and understanding situations in real time. Moreover, vision-guided robots may be largely calibration-free, which is a great practical advantage. Three vision-guided robots and their design concepts are introduced: an autonomous indoor vehicle, a calibration free manipulator arm, and a humanoid service robot with an omnidirectional wheel base and two arms. Results obtained, and insights gained, in real-world experiments with them are presented. Researchers and developers can take it as a background information for their future works.

Keywords: Machine vision (MV), Intelligence robots, human service, Robot guidance

63. PaperID 31031694: Significant Approach for Detaining Unpromising Contestant for Mining Pattern Based on Profit (pp. 541-544)

Vijay Kumar Verma, Lord Krishna College of Technology Indore

Kanak Saxena, Ph.D., Samrat Ashok Technological Institute, Vidisha M.P

Abstract - Today's every business organization needs profit. Professionals might give attention on recognizing its most treasured consumers who give a major portion of the profits to the business. Frequency based mining of items do not fulfill all the requirements of business. They only provide the information that an item has high low frequency based on a given value. There is one important factor profit has to be considered by every business. In past year a lot of methods have been developed for mining profit based pattern but efficiency, accuracy and scalability are important factors that have always to be considered. In this paper we proposed a significant approach for detaining unpromising contestant for mining profit based pattern. The proposed approach mines profit based pattern accurately and removes all unpromising contestants at different levels.

Keywords: Profit, Pattern, unpromising, frequency, efficiency

64. PaperID 31031676: Scalable and Secure Network Storage in Cloud Computing (pp. 545-551)

Muhib Ahmad Khan, M. Munwar Iqbal, Fahad Ubaid, Rashid Amin, Asima Ismail

Department of Computer Science, University of Engineering and Technology Taxila Pakistan

Abstract - Cloud Computing is a newly born type of computation, which depends on the shared resources of the network. Cloud Computing term discovered from that time when the system can access the different types of applications as well as different types of services remotely. Cloud Computing is the unique, next generation of IT architecture, in which computation is done on the open network shared resources, which create a security risk. In comparison to the existing conventional infrastructure, the IT services come under the IT expert control. In a market there is a different type of service provider using cloud computing features offers many different services like virtualization, applications, servers, data sharing, and try to reduce client-side computation overhead. Nevertheless, most of these services are outsourced to the third party, which creates the risk of data confidentiality as well as the data integrity. These days cloud computing, and its security is the hot topic for the research. In this paper, a new model is proposed for storage data on the network for the secure data storage on the cloud server, which achieves the security, availability, confidentiality and integrity.

Keywords - Cloud Computing, Data Integrity & Security, Data Confidentiality & Availability

65. PaperID 310316139: HSAG-FC: Hybrid of Simulated Annealing with Genetic algorithm and Fuzzy Classification for IDS (pp. 552-561)

Mrs. J. Lekha, Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women & Sri Krishna Arts and Science College, Coimbatore, Tamil Nadu

Dr. G. Padmavathi, Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, Tamil Nadu

Abstract - Nowadays signature attacks are termed as very big problem because it leads to software vulnerability. Malware writers confuse their malicious code to malicious code detectors such as Signature-based detection. However, it fails to detect new malware. This research article addresses the signature based intrusion detection from Intrusion Detection (IDS) systems. The proposed hybrid techniques for Generation of Signature are done using Genetic Algorithm (GA) and Simulated Annealing (SA) approaches. For this, signature-set in execution statements are selected by using simulated annealing and genetic algorithm, which produce the optimal solution of selection. Then the generated signatures are matched with IDS by using the two pattern matching techniques, namely (i). Finite state automaton based search for Single Pattern matching technique and (ii) Rabin Karp string search algorithm for multiple pattern matching technique. These techniques are used to match the signature as in an effective manner. In addition to this the Fuzzy Logic classification is used to find the degrees of truth of vulnerability for classification. The aim of the proposed work is to improve the final resultant accuracy in compared to existing techniques. The proposed Rabin Karp- fuzzy logic system returns the higher performance metrics namely precision is 88% and Recall is 80% and in open source dataset it contains 30 vulnerabilities this proposed worked well in detecting 28 vulnerabilities/ defect, the accuracy of this proposed is 94.27%.

Keywords: Degrees of truth, Finite state automaton, Fuzzy logic, Genetic algorithms, Intrusion Detection (IDS) systems, Optimization, Signature Generation, Signature matching, Simulated Annealing, Traffic detection.

66. PaperID 310316110: A Novel Technique for Jobs Scheduling In Cloud Computing Systems (pp. 562-568)

Muneer Bani Yassein, Yaser Khamayseh, Ali Hatamleh

Department of Computer Science, Jordan University of Science and Technology, Irbid, Jordan

Abstract — Recently, cloud computing has occupied a large place in the world, especially in the field of information technology. It is characterized as mainly rely on the Internet to provide services for organizations and consumers and to take advantage of resource sharing, in addition to that it is associated with many of the central remote servers to maintain user data, so it has become an effective way that will allow the world to use the many kind of applications without making an effort to be downloaded. Many job scheduling algorithms have been proposed to achieve both customer satisfaction and high resource utilization. However, better algorithms to achieve these goals efficiently are still needed. This paper proposes a hybrid technique for jobs scheduling based on Neural Network (NN) algorithm. The proposed algorithm classifies the jobs into four different classes. Furthermore, a Heuristic Resource Borrowing Scheme (HRBS) is proposed to exploit all services which has offered by cloud computing. Simulation is conducted using extensive (Cloud-Sim) simulator to measure the efficiency of the suggested algorithm in terms of average throughput, average turnaround time and average of context switch. The obtained results show that the proposed scheme outperforms other state of the art scheduling schemes.

Keywords - Cloud Computing, Job Scheduling, Hybrid Technique, Virtualization.

67. PaperID 310316116: Residual Energy based One-Hop Data Gathering in Wireless Sensor Networks (pp. 569-574)

Gaurav Kumar Pandey, Dept. of Computer Science and Engineering, Lovely Professional University, Jalandhar, India

Amritpal Singh, Dept. of Computer Science and Engineering, Lovely Professional University, Jalandhar, India

Abstract — The key constraint which hampers the performance of Wireless Sensor Networks is the limited battery power of the sensor nodes. Nodes once deployed cannot be recharged therefore data gathering from the sensor field should be done in such a manner that the energy of sensor nodes can be saved. Multi Hop routing and data relay protocols tend to deplete the battery power of the forwarding nodes at a large extent. Also, Clustering Algorithms generate extra overhead which affects the lifetime and performance of the network. In this paper we introduce Residual Energy based One-Hop Data Gathering (REO-HDG) in Wireless Sensor Networks by making use of a Mobile Data Collector (MDC) that traverses the sensor field and collects data from the sensors using single hop only, which in turn eliminates the problem of data relay. We make use of rendezvous locations, one-hop neighbor sets and residual energy of sensors to gather data from the sensor nodes. The union of all neighbor sets include all the candidate sensor nodes. REO-HDG tends to maximize the lifetime of the sensor network by eliminating data relay and clustering.

Index Terms— *Mobile Data Collector (MDC), Data gathering, Residual Energy, Energy Conservation, MDC Scheduling, Wireless Sensor Networks.*

The Smartcane for Blind People An Electronically Smart Stick to Aid Mobility

M. Asad Hussain^{1*}, M. Ghazanfar Ullah¹, Atif Fareed¹, Basit Sohail

¹Department of Electrical Engineering, Usman Institute of Technology – Pakistan

Abstract—This paper is focused about the development of a “Micro-controller based Smart White Cane” A.K.A. “The Smartcane” and its comparison, based on performance and usability, with other existing models. Our main contribution is to enhance the capabilities of existing models of micro-controller based white stick for blind persons, due to their practical limitations. The developed project serves the best solution to overcome the difficulties of blind people, so that they can easily mobilize themselves, be a more successful part of society. The developed project facilitates blind persons in a manner that they can handle any obstacle, wet material, uneven surface, etc. Our main objective was to reduce the size of the presented model by integrating the circuits and making it a compact and portable stick for users. Also, we emphasize on the range of the modules and sensors to increase the efficiency and usability of the prototype model. The system accompanied a portable unit that can easily be carried and operated by a visually impaired user. It could easily be incorporated into a walking cane. The salient features of the developed prototype are ultrasonic sensor for obstacle detection, water probe for mud and water detection, I.R. for ditch detection, G.P.S, G.S.M. module, signal-to-speech module, speaker or headset, and portability (size and power). The experimental results shows that the developed prototype is much more efficient and usable in varying situations for a blind person as compared to the ordinary white sticks while affordable and cost effective at the same time.

Keywords – Blind, Mobility aid, Smartcane, Microcontroller, GPS, GSM, Ultrasonic sensor, IR sensor.

I. INTRODUCTION

In today’s advanced technological world, the need of autonomous living is undisputed. The main problem of social exclusiveness is seen in case of visually impaired people. They suffer in an unknown environment without any manual assistance which is difficult to get at all the time. According to WHO (World Health Organization), 285 million people are estimated to be visually impaired worldwide in which 39 million are blind and 246 million have low vision [1]. About 90% of these people live in developing countries. An inability to interact with the environment due to blindness becomes a real challenge for most of them, although they rely on their other senses. To assist visually impaired person, traditionally a white cane commonly known as walking cane is used, a simple mechanical device to detect the ground, uneven surfaces, holes and stairs using simple tactile-force feedback. Although such person rely on other senses, but, walking through an unknown environment becomes a challenge for most of them. However the device considerably fails in case of dynamic obstacles owing to the noise they produce. [2]

Currently most blind people rely on other people, dogs, and their canes to find their way in buildings. This can be a hassle for both the visually impaired person as well as others. Many disabled people prefer to do things independently rather than rely on others. The Smart Blind Stick can provide a solution to this problem.

Main objective concern is the enhancement of microcontroller based stick to facilitate the disable community. This project provides the best solution for the difficulties of the blind people, by which they can easily, mobilize themselves, be successful part of society, earn their living in easy manner and get a position which suits them positively. This project facilitates blind in a manner that they can handle any obstacle, slippery material, uneven surface, etc. Our main objective is to reduce the size of the present model by integrating the circuits and making it a compact and portable stick for users. Also, emphasizing on the range of the modules and sensors to increase the efficiency and performance of the model. The developed prototype intends to provide a portable unit that can easily be carried and operated by a visually impaired user. It could easily be incorporated into a walking cane [3].

II. PROBLEM STATEMENT

Providing facilities to disable persons should be our priority and social obligation. As blind persons suffer a lot in their daily routine work and become much more dependent on others to manage their works properly. The developed prototype is a “Real Model of Microcontroller Based White Cane”, AKA “The Whitecane”, which can help visually disabled persons in their daily work so as they get interact with the environment without facing any problem. It is our aim to provide the aid and hassles free mobility to the Blind Person. This project provides the best solution for the mobility difficulties of the blind people, by which they can easily, mobilized themselves, be successful part of society, earn their living in easy manner and get a position which suits them positively. This project facilitates blind’s in a manner that they can handle any obstacle, wet material, uneven surface, etc. [4].

III. RELATED WORK

From the decades, several researches and developments have been made to design and implement the new devices to aid visually impaired community. By the advent of time various devices have been made and still improving every passing day. These works are mainly focused on three types of environments: outdoor, indoor, and some mix of the previous ones [5].

Shoval et. al in [6] developed at Navbelt, a portable computer to avoid obstacle that can only be used for indoor navigation. The computer can be operated in two modes. The first mode is used for conversion of system information into sound signal. This produces two sounds, one indicates the direction of travelling and other indicates the blockage of passage. It is hard for user to differentiate between the sounds. Also, system was incapable of determining the momentary position of the user.

D. Yuan et al. in [7] have made a discussion regarding the virtual cane sensing equipment that can able to measure the distances with the measuring rate of 15 measurements/second. This device is used as the flash light. The device can also detect uneven surfaces by analyzing range of data collected as the device swings around. Przemyslaw Baranski et al. in [8] have discussed the concept of remote guidance system for blind person. The system is divided into two major parts – one is the operator’s terminal and the other one is the mobile terminal for blind. This mobile terminal contains a digital camera, GPS and headset.

Both the terminals are wirelessly connected through GSM and internet. The link is responsible for transmitting the video from blind traveler, GPS data and providing audio communication in between the terminals.

Rupali Kale and A. P. Phatale in [9] have designed a fully automated system to aid the blind person in all respects. The system comprises on GPS and object prevention technologies.

Sabarish.S in [10] has discussed the idea for the development of the navigation system for the assistance of blind person to move around without any problem. The system is based on microcontroller technology with speech module. Also, it contains vibrators, ultrasonic sensors mounted on cane as well as on the shoulders of blind person.

V.Dhilip Kanna et al. in [11] presented a virtual eye system consists of PGA and detectors. The virtual eye in this project is a small camera that communicates with the outside environment and it is a constant source of information to the blind person.

Ranu Dixit and Navdeep Kaur in [12] proposed a system that is based on HMM algorithm. The system used to read different sound signals bit wise. The signal that are read stored in database where the HMM algorithm is applied on the signal and every signal is used as one single information.

WaiLun Khoo et al. in [13] have made the discussion that how one can utilize the use of wearable range-vibrotactile device in real world environment. Also this is demonstrated using simulations in virtual environment for the accomplishment of complicated navigation tasks and neurosciences.

IV. DESIGN

The design process is based on the architecture illustrated in figure 1. Essentially, “The Smartcane” functions like ordinary blind canes. The difference is that the Smartcane is equipped with ultrasonic sensor, water sensor, GPS and GSM modules as illustrated in figure 1. Also the cane is designed to be foldable so that it is easy for the user to keep and handle [14].

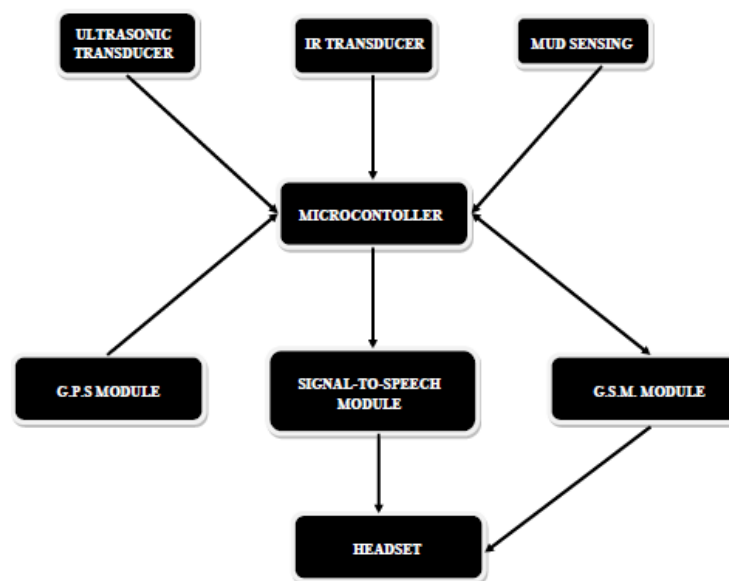


Figure 1: Block diagram of the system

A. Hardware Design

Initially the design of the stick was modeled on CAD software. The major constraint while designing the hardware is that it must be light weight, easily to operate without any problem and provides full functionality to blind person. By analyzing, it was concluded for the design that all the sensors should be mounted on the stick and all the hardware circuitry should be embedded in a belt pack that is clipped on the waist of the user as illustrated in fig.2. So the hardware design is then divided in two major parts. One is the stick and the other one is the belt pack. The sensor and circuitry is connected with the help of connector [10].

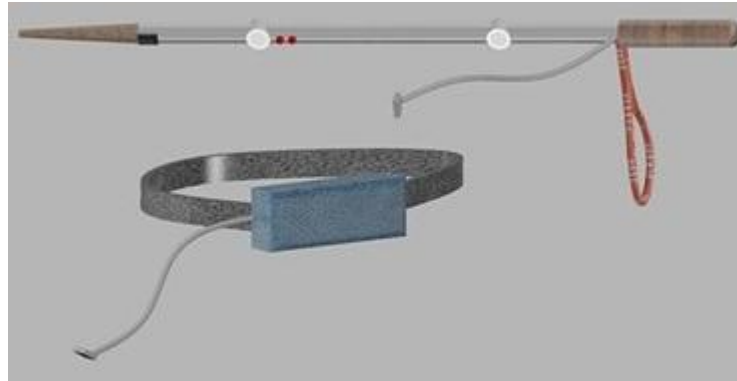


Figure 2: The Smartcane stick and belt modules

1) *Stick*: The stick is 3'ft long, 0.75" inch diameter, carbon fiber body with ergonomic handle and a wire strap of rubber. There are three types of sensor mounted on the stick. First is ultrasonic sensor. Second is the Infrared (I.R.) sensor and the third one is the mud sensor. The ultrasonic sensor is placed on the height of 1 ft. and the I.R. sensor at height of 0.5ft and at the base of the stick is mud sensor.

2) *Belt Pack*: The belt pack contains all the circuitry that includes main controlling circuit, sensor circuits, GSM module, GPS module, Speech module and speaker along with headset. The circuitry and sensor are connected via connector as illustrated in the figure 2.

B. Circuit Design

For various modules, few circuits were designed. First, the proto-board was used for design test and for easy modification task. Once the circuit on the proto-board was finalized, printed circuit board (PCB) was fabricated. The size of the PCB was considered as an important factor in circuit designing; which is long rather than wide, as



Figure 3: Complete Circuit Board

illustrated in figure 3. It is also to make sure that the design of The Smartcane is as small as possible. The whole circuitry is fixed in a belt pack that is associated with the stick through wires. The stick only contains the sensors.

1) *Main Board*: As illustrated in figure 4, the main circuit board comprises on microcontroller. The microcontroller used for this project is AT89C52. This microcontroller is a low power, high performance 8-bit

microcontroller with CMOS technology. It has built-in 8kB PEROM. The main reason behind selecting this particular microcontroller is its greater flexibility and cost effective solution [15].

All of the circuitry and sensors are centrally connected to the microcontroller via the main board. All the controlling is also done through this board.



Figure 4: Main Circuit Board

2) *Ultrasonic Sensor*: Ultrasonic Sensor Module consists of ultrasonic sensor along with the circuitry integrated on the same board. They work on the principle alike to radar that senses for the target by inferring the echo signals [16], as presented in figure 5. The Ultrasonic Sensor module that is used in this project is HC-SR04.

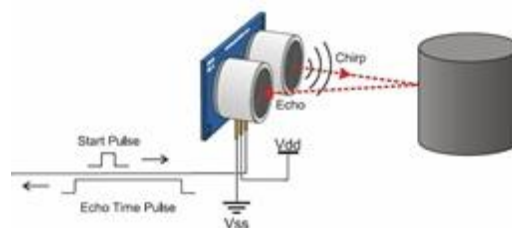


Figure 5: Working of Ultrasonic Sensors.

3) *Infrared Sensor*: The infrared sensor that is used in this project is adjustable that can be able to detect the distance ranging from 3cm to 50cm. it is used for the detection of any ditch or even surface [17]. This sensor is shown in the figure 6 along with circuit diagram in figure 7.

4) *Mud Sensor*: The mud detector module is constructed using PC817 Optocoupler IC that gives the signal when the input pins are shorted by any conducting material.



Figure 6: Infrared Sensor

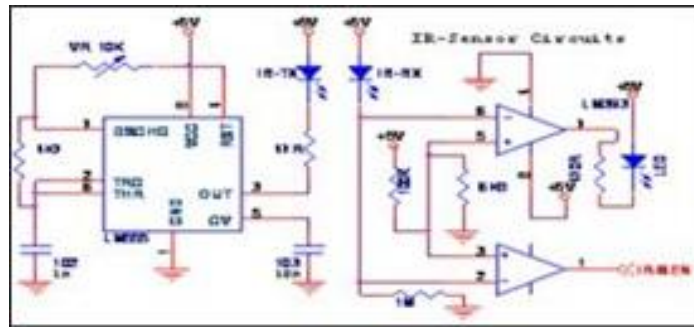


Figure 7: Circuit Diagram for Infrared Sensor Module.

5) *Signal-to-speech Module*: Referring to figure 8; the signal-to-speech module is one of the most distinctive and important feature of this project. This module contains pre-recorded voice messages that activates on every detection of any sensor. Also it activates when emergency or panic push buttons are pressed by the user. This module contains a speaker as well as jack for audio out where earphones can be connected to listen the prerecorded signals [10].

The voice that we have used in our project is ISD2590. This chip is capable of providing high quality record and playback solutions for messaging applications of 60 to 120 seconds. The main purpose of using this chip is its automatic gain control, smoothing filter, speaker amplifier, and anti-aliasing filter and contains high density multi-level storage array [18].

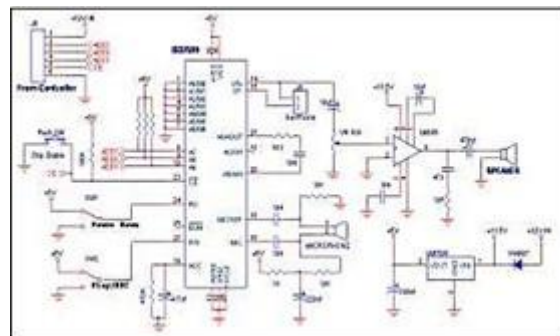


Figure 8: Circuit Diagram for Signal-to-speech Module

6) *GSM/GPS module*: The GSM Module used in this project is Blue Ocean GSM-S-A2. This module, as shown in figure 9, is suitable for SMS, data and fax applications. This is used for messaging and is integrated via system using serial port RS232. This GSM is light in weight and easy to integrate with AT89C52 microcontroller [19].



figure 9: GSM/GPS Module. [19]

7) *Battery Circuit*: The system is powered up by 12V rechargeable battery circuit, as described in fig. 10. The power requirement for the system is very low as no such heavily powered equipment is used.

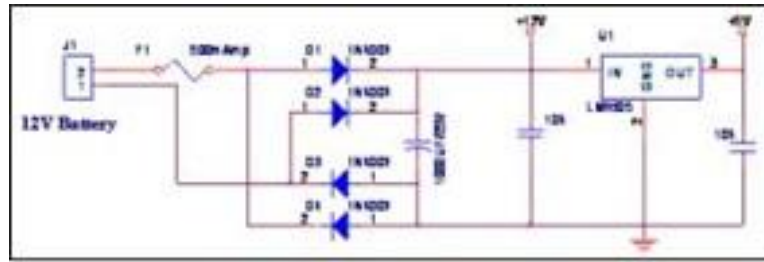


Figure 10: Circuit Diagram for Battery

V. WORKING

The stick can efficiently be used as an electronic guide to the blind person. A blind person can use it easily during walking on the road or moving in any building. The whole system is powered by a 12 V rechargeable battery. The implemented I.R. Sensors are responsible to detect any uneven surface or any ditch ahead of blind person. Furthermore, the Ultrasonic sensors are fabricated in the stick to detect any wall or obstacle in front of blind person. The function of probes is as a mud detector or we can say that it is use in the stick as a water or slippery material detector to save user from any injury or serious hurt. The voice card is also associated in the circuitry that send the prerecorded messages over the ear phone and speaker attached to the voice card so that the blind person can be alerted from any detection. The G.P.S. module is provided for location tracking of the blind person. If any other person wants to know the location of the user, he simply sends a coded text over the SIM-card that is being placed in the GSM module. The GSM will send back the last saved coordinates to that particular number. Moreover different push buttons are provided in the stick. In case of any emergency, a text message is being generated to the number that is being fed in the memory by pressing the push buttons [20]. The working process of the developed prototype is illustrated through figure 11.

VI. RESULTS AND DISCUSSION

The developed prototype was tested on 10 visually impaired people (7 in between the age of 20 to 50, and 3 within the age of 10 years) and 20 people who could see but have blindfolds on their eyes. All were tested by offering same pathway (200 meters length outdoor), containing obstacles of various types, to walk. They were evaluated on the basis of the rate of collision, walking speed and the usability on how they interact with the environment using our developed prototype “The Smartcane” and with the “ordinary white stick”. A comprehensive performance comparison has been presented in table 1, 2 and 3. The developed prototype has also proved to be in reasonable price and weight. A comparison of the developed prototype “The Smartcane” with other existing similar products [21-23] in the market has been presented in figure 12 and 13.

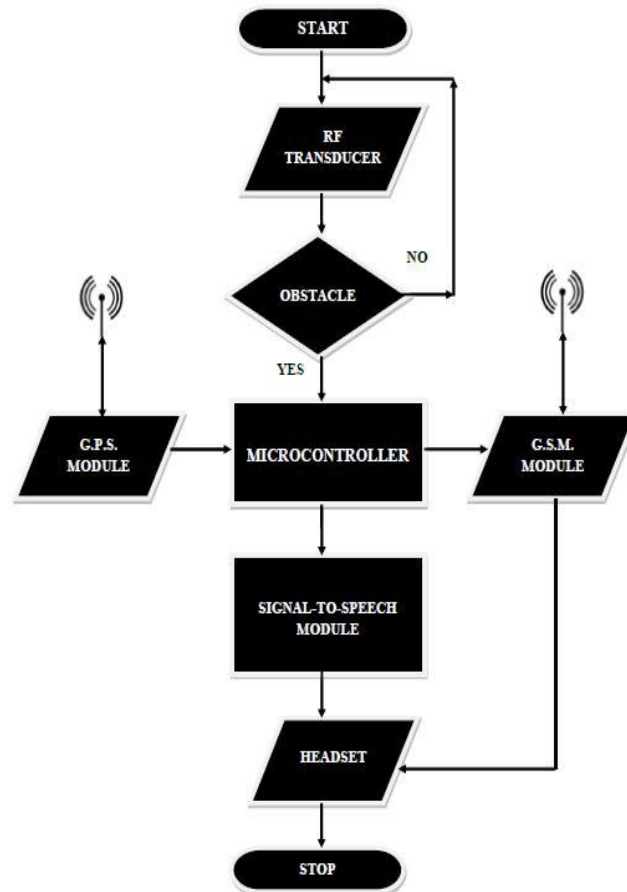


Figure 11: Process flow diagram of the developed prototype

Table 1: Performance test on blind people (10 in number)

Average collision rate (%) with obstacles after 200 meters outdoor walk		
Obstacle	The Smartcane	Ordinary White Stick
Above chest	98%	100%
Average chest height	85%	98%
Above waist and below chest	25%	50%
Below waist	1%	15%
Mud Detection	10%	100%
Wet garbage	40%	90%
Dry garbage	70%	70%
Wet surfaces	10%	90%
Uneven surface	8%	25%

Table 2: Performance test on blindfolded people (20 in number)

Average collision rate (%) with obstacles after 200 meters outdoor walk		
Obstacle	The Smartcane	Ordinary White Stick
Above chest	90%	100%
Average chest height	88%	100%
Above waist and below chest	30%	65%
Below waist	7%	24%
Mud Detection	8%	100%
Wet garbage	42%	95%
Dry garbage	80%	85%
Wet surfaces	25%	95%
Uneven surface	10%	40%

Table 3: Average walking speed over a distance of 200 meters

	Average walking speed (m/s)	
	With "The Smartcane"	With Ordinary White Stick
Blind persons	1.1	0.8
Blindfolded persons	0.7	0.5

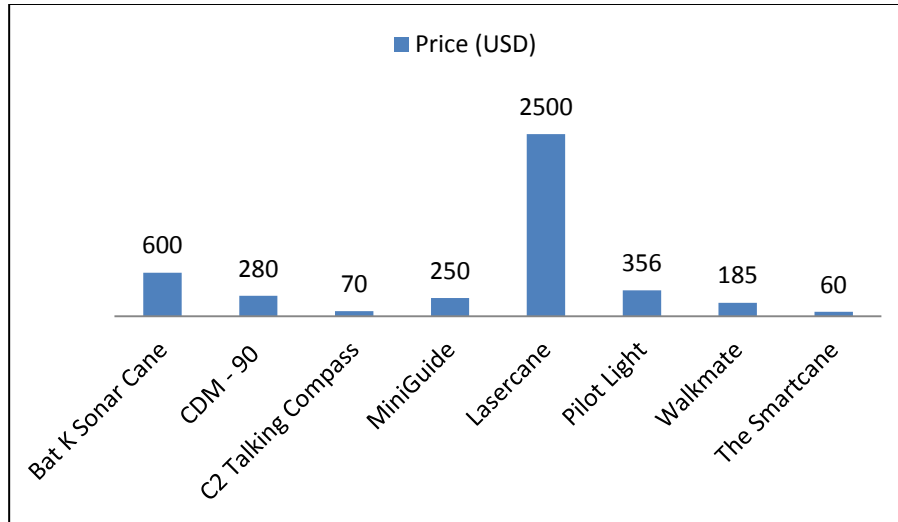


Figure 12: Price comparison of "The Smartcane" with similar market products

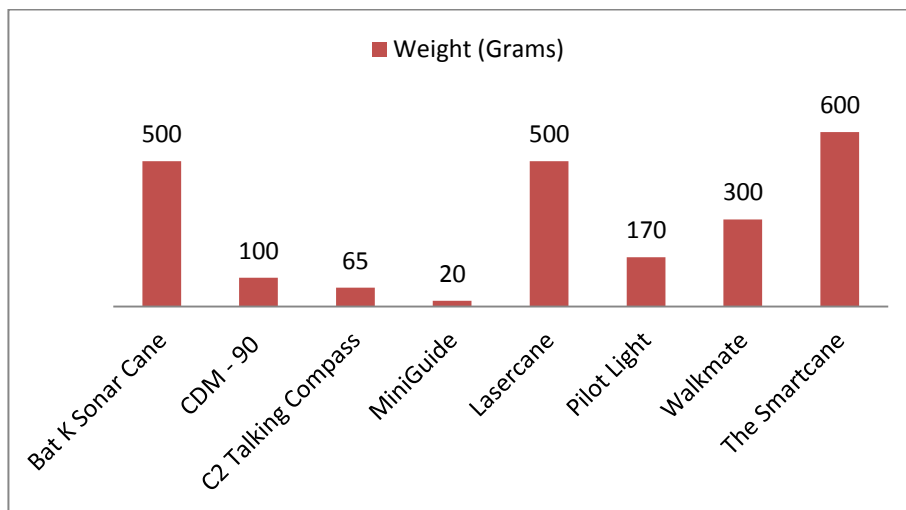


Figure 13: Weight comparison of "The Smartcane" with similar market products

VI. CONCLUSION

After rigorous testing in varying situations, this project proved to be completely an electronically smart system that provides the blind and visually impaired persons to live their lives without the help of others. The system is capable of providing smart assistance to the blind person in a manner that it can detect any obstacle, any uneven surface and provide aid in the case of emergency. This system is practically tested on blind persons as well as blindfolded persons and as a result they feel very comfortable in operating the system. Furthermore, as evident through figure 13, there is a need to further reduce the weight of the system, however, this issue is also addressed intelligently as the battery and circuit module can be tied-up around the waist, hence, the weight of the stick is reduced significantly.

REFERENCES

- [1] World Health Organization, "<http://www.who.int/mediacentre/factsheets/fs282/en/>", last accessed February 2, 2016.
- [2] Hiebert J., Jessani F., et. al., "Intelligent Cane", Proceedings of ENGG 3100: Design III projects, University of Guelph, pp: 21~27, 2007.
- [3] Abd Wahab M. H., Talib A.A., et.al, "Smart Cane: Assistive Cane for Visually-impaired people", International Journal of Computer Sciences Issues (IJCSI), Vol. 8, issue 4, no. 2, July 2011.
- [4] Sharma P., Shimi S. L., Chatterji S., "Design of Microcontroller Based Virtual Eye for the Blind", International Journal of Scientific Research Engineering and Technology (IJSRET), Vol. 3, Issue 8, pp: 1137~1142, November 2014.
- [5] Luis A. Guerrero, Francisco Vasquez and Sergio F. Ochoa, "An Indoor Navigation System for the Visually Impaired", Sensors 2012, vol. 12, pp. 8236-8258; doi:10.3390/s120608236
- [6] Shruga Shovel, Iwan Ulrich, and Johann Borenstien, "NavBelt and the Guide Cane", IEEE Transactions on Robotics & Automation, Vol.10, No.1, pp. 9-20, March 2003.
- [7] D. Yuan and R. Manduchi, "Dynamic Environment Exploration Using a Virtual White Cane", Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), University of California, Santa Cruz, pp. 1-7, 2005.
- [8] Przemyslaw Baranski, Maciej Polanczyk and Pawel Strumillo, "A Remote Guidance System for the Blind", IEEE Transactions on Remote Guidance, pp. 386-390, June 2010.
- [9] Rupali Kale and A. P.Phatale, "Design of a GPS based Virtual Eye for the Blind People", International Journal of Current Engineering and Technology, Vol.4, No.3, pp. 2162-2164, June 2014.
- [10] Sabarish.S, "Navigation Tool for Visually Challenged using Microcontroller", International Journal of Engineering and Advanced Technology (IJEAT), Volume-2, Issue-4, pp. 139-142, April 2013.
- [11] V.Dhilip Kanna, P.V.Deepak Govind Prasad, S.Aswin Amirtharaj and N.Sriram Prabhu, "Design of a FPGA based Virtual Eye for the Blind", Proceeding of Second International Conference on Environmental Science and Technology, Singapore, Vol.6, No.2, pp. 198-202, 2011.
- [12] Ranu Dixit and Navdeep Kaur, "Real Time Information System Based On Speech Processing: Radio Station As Virtual Eye For The Blind", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 2, No. 2, pp. 362-371, February 2013.
- [13] WaiLun Khoo, Joey Knapp, Franklin Palmer, Tony Ro and Zhigang Zhu, "Designing and Testing Wearable Range Vibrotactile Devices", International Journal of Assistive Technologies, Emerald Group Publishing Limited, Vol.7 No.2, pp. 102- 117, 2013.
- [14] Muhammad N., Ali Q. W., "Design of Intelligent Stick Based on Microcontroller with GPS using speech IC", International Journal of Electrical and Computer Engineering (IJECE), Vol. 2, No. 6, pp: 781~784, December 2012.
- [15] Atmel Corporation, AT89C52 8-bit Microcontroller with 8K Bytes Flash Data Sheet. Atmel Corporation 1999, <<http://www.atmel.com/Images/doc0313.pdf>> Accessed Sept. 15, 2015.
- [16] Paradkar P. K., "Ultrasonic Sensor HC-SR04 Interfacing with 8051 Microcontroller", <http://robotronicspro.blogspot.in/2014/11/ultrasonic-sensor-hc-sr04-interfacing.html#.VfesLUBfzic>, last accessed on Sept. 15, 2015.
- [17] "Adjustable Infrared Sensor Switch SKU: SEN0164", http://www.dfrobot.com/wiki/index.php/Adjustable_Infrared_Sensor_Switch_SKU:SEN0164, last modified on Sept. 25, 2013, last accessed on Sept. 1, 2015.
- [18] Winbond Electronics Corporation, "ISD5260/75/90/120 Single-Chip, Multiple-Messages, Voice Record/Playback Device 60-, 75-, 90- and 120- Second Duration", <<http://www.futurlec.com/Others/ISD2590.shtml>>, May 2003, last accessed on Sept. 7, 2015.
- [19] BluOcean, "SMS software for BluOcean GSM-S-A2, GSM-S-U2, GSM-S-W2 GSM GPRS modem series", <<http://www.activexperts.com/hardware/gsm-modems/bluocangsms/>>, last accessed on Aug. 18, 2015.
- [20] Lohani K., Sakhuja A., Butola R., "Techno-Cane – Guide for the Visually Impaired", International Journal of Engineering Technology, Management and Applied Sciences (IJETMAS), Vol. 3, Issue 6, pp: 1~4, June 2015.
- [21] <http://www.sensorytools.com/c2.htm>, last accessed July 20, 2015.
- [22] http://www.gdp-research.com.au/minig_1.htm, last accessed July 20, 2015.
- [23] <http://www.noogenesis.com/eta/current.html>, last accessed July 20, 2015.

E-Commerce Framework Based on Evaluation of Data Mining and Cloud Computing

Mohd Muntjir

College of Computers and Information Technology, Taif University
Taif, Saudi Arabia

Ahmad Tasnim Siddiqui (Corresponding author)

College of Computers and Information Technology, Taif University
Taif, Saudi Arabia

Abstract - This paper is a description about the application of e-commerce and data mining with cloud Computing. It emphasizes how data mining is used for e-commerce in combination of cloud computing systems. Data Mining is a process of separating possibly useful information from available raw data. It's also describing that How SaaS is very useful in cloud computing. The combination of data mining techniques into normal day-to-day actions has become common part. Businesses and advertising have become more active through the use of data mining functionalities to deduct the overall costs. Data mining operations can develop much more demographic information respecting customers that was basically not known or hidden in the desired data. It has basically seen enhancements in data mining techniques proposed to such activities as identifying criminal activities, fraud detection, suspects, and indication of potential terrorists. On the whole, data mining systems that have been designed and developed to data for grids, clusters, and distributed clusters have considered that the processors are the limited resource, and hence distributed. When processors become accessible, the data is transferred to the processors.

Keywords: Data Mining, e-commerce, cloud computing systems, data mining and cloud computing, (SaaS) Software-as-a-Service.

I. INTRODUCTION

Data mining is the abstraction of invisible guessing information from huge databases, is a strong new technology with great latent to help companies focus on the most important information in their data warehouses. Data mining tools presume future methods and trends with its behaviors, granting enterprises to make intense, knowledge-driven outcomes [1]. The motorized, proposed analyses offered by data mining move before the analyses of past events implemented by recollected tools typical of decision support structure. As data sets have increase in size and complication, towards hands-on data analysis has increasingly been expanded with ambiguous, automatic data transforming. It's been assisted by other explorations in computer science, such as genetic algorithms (1950s), neural networks, decision trees (1960s), support vector machines (1990s) and cluster analysis. Data mining is the technique of applying these methods to data with the intention of uncovering invisible patterns in big data sets. Furthermore, Data mining is categorizing through data to considering patterns and implement relationships [6].

1.1 Framework about data mining

1. Association –It means, looking for connections where one event is associated to other event.
2. Sequence or path analysis – It analysis for arrangements, where one event leads to another event
3. Classification – It distributes some new patterns
4. Clustering – It finds and visualize, documenting groups of facts not before known
5. Forecasting -It discovers patterns in data that can lead to acceptable predictions about the future analysis. Forecasting part of data mining is considered as predictive analytics.



Fig 1: -Data Mining Framework (Source: -<http://www.sqlservercentral.com/blogs/zoras-sql-tips/2014/10/15/an-introduction-to-sql-server-data-mining-algorithms/>)

For instance Visual Numeric's has been delivering latest foretelling and data mining results across a huge range of enterprise and trades such as financial services, healthcare, aerospace, government and telecommunications,. Visual Numeric's' forecasting solutions integrate technical skills, decades of hands-on experience and powerful products to create the highest quality solutions possible for your visual data analysis requirements. Alike, so there are different applications of data mining in real world as, Data Integrity, Space Organization, Hospital, Airline Reservation, Student Management, Forecasting, Biometrics, Web Mining, Parallel Processing, Geographical, Mathematics, and many more services [5]. Data mining uses complicated mathematical algorithms to section the data and assess the possibility of future proceedings. Data mining is also recognized as Knowledge Discovery in Data (KDD) [11].

II. DESCRIPTION ABOUT CLOUD COMPUTING

Cloud computing is a basic term for anything that integrates distributing hosted services over the Networks. These services are basically distributed into three orders: Infrastructure-as-a-Service (IaaS), Software-as-a-Service (SaaS), and Platform-as-a-Service (PaaS). Regarding the name of cloud computing, it was inspired by the cloud symbol that's always used to express the Internet in diagrams and flowcharts [2].

The real term "cloud" obtains from telephony in that telecommunications companies, who until the 1990s proposed initially committed point-to-point data circuits, began offering Virtual Private Network (VPN) relevance with desirable quality of service but at a reasonable cost. Early in 2008, Eucalyptus became the first open-source, AWS API-appropriate platform for extending private cloud. In early 2008, Open Nebula, increased in the RESERVOIR European Commission-funded project has become the first open-source software for deploying local and hybrid clouds and for the federation of clouds.

In June 2, 2008 - Cloud computing is becoming one of the further industry latest word. It integrates the ranks of words including: utility computing, grid computing, clustering, virtualization, etc. Cloud computing imbricates some of the concepts of distributed, grid and utility computing system, although it does have its own meaning if it is

contextually used correctly. The conceptual overlay is basically due to technology usages, changes and implementations over the many years. The cloud is a virtualization of resources that manages and sustains itself. Off course, there are people reserves to keep operating systems hardware and networking in proper orders. Although from the context of a user or application developer only the cloud is accredited [4].

Cloud computing actually is collecting resources and services required to perform operations with strategically changing basic requirements. A service developer or an application, grants access from the cloud rather than a particular endpoint or named source. As because of the fast progress in a network technology, the cost of broadcasting a terabyte of data over long distances has reduced extremely in some past decade. The entire cost of data management is five to ten times higher than the basic receiving cost. According to all consequences, there is an enlarging interest in outsourcing database management employments to third parties that can arrange all of these jobs for lower cost due to the savings of scale [12].

2.1 Architecture provides structures

- Self-monitoring
- Self-healing
- Resource registration and discovery
- Automatic reconfiguration
- Service level agreement definitions

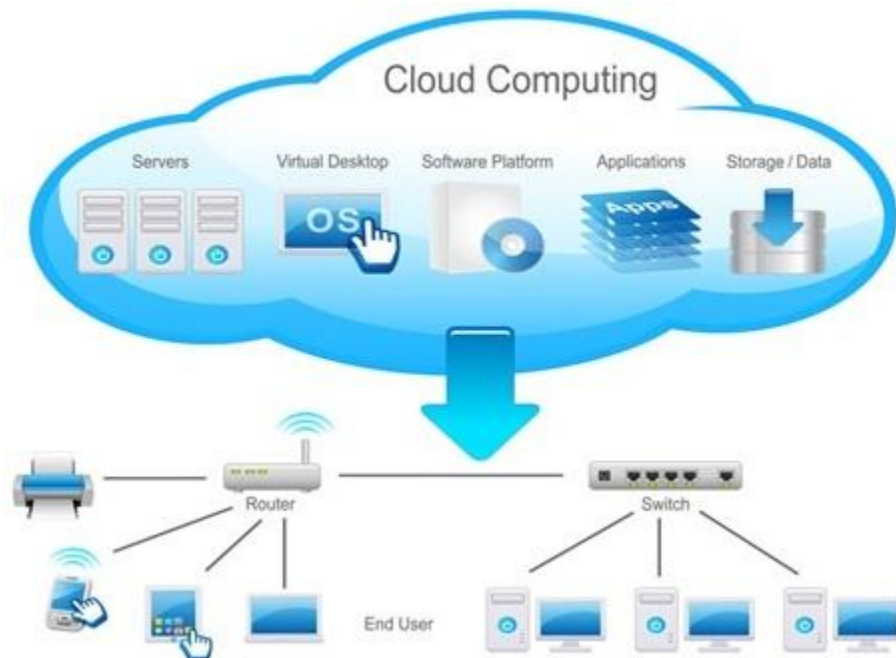


Figure 2: Cloud Computing Sample Architecture
(Source: <http://rubiconn.com/services/cloud-computing/>)

III. SPOTLIGHT ON CLOUD SERVICE

Cloud has three types of services, platform as a Service, infrastructure as a Service, Software as a Service. In which SaaS is king of all the services.

PaaS:

- Provides a platform or solution stack on cloud systems.
- Offers browser based development atmosphere.
- Integrates built-in security, scalability and web service interfaces.
- Provides web service interfaces that privileges us to connect the applications outside the platform
- Provides a facility for Lower cost and improved profitability
- Sits on a top of the IaaS architecture and combines with progression and middleware abilities as well as messaging, database, and queuing functions.

IaaS:

- Provides On-demand self-service
- Provides capabilities of broad network access over the internet.
- Provides capabilities of automatically control and optimize the resources in a cloud systems
- Provides computer atmosphere as a relevance service, basically in a virtualized environment.
- Provides dynamic scaling, policy based services and desktop virtualization
- Delivers numerous capabilities for adaptability and measurement.

SaaS:

- Provides the application over the Internet or Intranet via a cloud Framework.
- Provides service for an automatic updates and patch management
- Provides easy administration and global accessibility.
- In SaaS, Application Programming Interfaces allow for integration between different types of software
- Provides a facility for software delivered in a “One-to-Many” model
- Constructed on fundamental IaaS and PaaS Layers.



Figure 3: Layers of Cloud Computing
(Source: <http://www.cloudsymposium.com/importance-of-cloud-technology/categories/>)

IV. DECREASE DATA MINING EXPENSES BY SAAS – CLOUD MINING IS BORN

SaaS Distribution model (Software-as-a-Service) provides to decrease values by giving elastic license options and outsourcing the hardware endeavor. SaaS Solution basically provides billions of records each month by using the power of the cloud to maintain latest analytics to recognize more recovery contingency. One of the fundamental profits of the SaaS model, this scale up/scale down efficiency makes sure to pay only for what is used and use only what you need in an easy approval model.

At Software-as-a-service (SaaS), software's are not used in the company; it comes at a software service provider's server side. It means the provider contacts with the hardware, follows software updates and manages everything in details. In Cloud Mining a server servers that maintain the software are the Cloud.

It can be the public cloud from Amazon.com, Google, etc., or a private cloud on the servers provided by a single provider or many providers. It has two main holdings; on one way the customer only remunerations for the tools of Data Mining he requires. It makes him save a lot related to complex Data Mining suites that he is not using extensive. And on the other way he just pays for the costs that are developed by using the Cloud Systems. It does not have to manage a hardware framework; he can implement data mining just via his web browser. This decreases the barriers that keep small businesses from improving of Data Mining.

4.1 Key Characteristics of SaaS

- Centralized feature updates: This prevents the requirements for downloadable patches and upgrades average of on-premise software installations.
- Single-instance, multi-tenant architecture: A one-to-many model indicates a single physical instance with customers hosted in separate logical space. There may be different variations of how a single instance really gets completed and how multi-tenancy really gets attained.
- Managed centrally and accessed over the Internet: Basically, there is no software element installed at the customer sites. Although, all applications can be secures remotely over the websites.
- Generally priced on a per-user basis: Minimum number of users that companies can sign up for varies from one SaaS vendor to another and also depends on what stage the SaaS vendor is in their expansion path as a company or business. Many of them do charge additional fees for extra bandwidth and more storage.
- Mostly subscription-based, no upfront license costs: It indicates that functional leaders (from marketing, sales, HR and manufacturing) do not have to go through their IT department to maintain and get them approval.

4.2 Key Drivers behind Adoption

- Improved network bandwidth.
- Security and safety are sufficiently well-trusted
- Reliability and popularity of web applications
- Low cost of ownership

V. CLOUD MINING BY LAYERED COMPUTER TECHNOLOGIES

In modern days, layered Technology is a leading global provider of managed and maintained dedicated hosting, Web services, and cloud computing / on-demand virtualization. By maintaining high-quality technology framework, infrastructure and support, Layered Tech implements customers to remove capital expenditures and save on operating costs while centralizing on focused business issue. The Layered Tech's extensible framework powers millions of sites and Internet-enabled applications including software as a service (SaaS), e-commerce, content

distribution and many more applications. Furthermore, our clients range from leading-edge Web 2.0 startups, strong mid-sized businesses and some of the world's biggest consultancy and alliance enterprises [3].

The Layered Technologies, a governing worldwide service provider of on-demand Information Technology frameworks, has designed a modern virtual private data center (VPDC) platform with levels of maintained services, all security and flexibility via a recovery API that were formerly inaccessible in a unified offering. A new platform is a hybrid cloud computing infrastructure that provides customers a virtualized environment on dedicated servers within Layered Tech data centers, along with levels of extensibility on how to safe access their VPDC, even if dedicated lines, VPN or Internet services.

Layered Tech's latest alliance virtualized platform provides customers the accuracy of dedicated servers with the high possibility, high processing power and scalability of virtual machines to meet regularly changing business needs with basic requirements. Customer API, designed by Layered Tech and based on industry standard protocols (SOAP and XML-RPC), arranges easy communications and enables customers to establish activities such as customizing proprietary applications framework, maintaining and managing resources, investigating analytics and more via computer or mobile devices. The service provider of on-demand IT infrastructure tries to swamp user's involvement if their data is secure by having a business level security standard. Furthermore, it will be possible to design, develop, order and set up a safe virtualized environment within a desired period. A customer can elect whatever platform they required, including VMware and Microsoft Hyper-V, 3Tera's AppLogic.

VI. E-COMMERCE AND IMPACT OF CLOUD COMPUTING

Cloud storage is the application of cloud computing and it is a type of facility provided by the cloud computing technology. Using the functionality with the cluster system, grid technology or distributed system, etc., cloud storage can cluster the various kinds of storage equipment inside network to cooperate together and gives external data storage and access service by application programmes.



Figure4: The E-commerce cloud

The cloud storage may provide personal and enterprise cloud storage services. Some common services of e-business include online store, customer management, e-stores/digital stores, invoicing & packaging, payment options and shipping [13]. These services can be used on cloud. The consumer or enterprise starting e-commerce business or e-business may rent the cloud storage for the e-commerce data storage. On the other hand, it may well

reduce an investment around the software and hardware; in contrast, additionally, it can reduce the price on taking care and operation. Besides, many of the cloud storages possess the remote disaster recovery and data backup system, which can ensure the safety on the e-commerce data at minimum affordable cost.

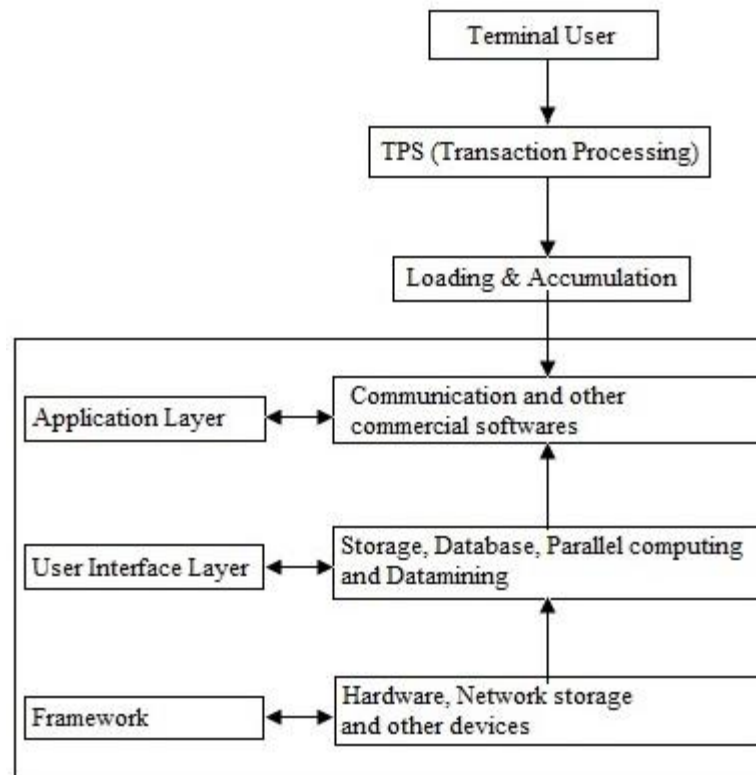


Figure 5: E-commerce framework for cloud computing and data mining

Cloud computing in e-commerce refers back to the policy to pay for bandwidth and storage area using a scale depending on the usage. This process differs from the more common method wherein anyone insures some drive space and bandwidth. Cloud computing in e-commerce makes really a computer program, on demand basis where anyone pays less in case you have less traffic visiting e-commerce sites. Due to drastic decline in the setup and maintenance costs in cloud computing, many businesses have started changing the manner and started switching from having corporate-owned hardware and software to presenting cloud style business models using pay per use models.

These days cloud computing is being promoted very significantly by e-commerce organizations as the money necessary for storing large number of business data might be reduced into a large extent by storing the details in cloud data centers. This technique of Hosting e-commerce systems provides both commercial information like products prices and available quantities and even facilitates various commercial actions like buying, selling and negotiation [10].The cloud computing as a brand new service model, with network storage, on-demand use of nature, provides a brand new information resource sharing and processing mechanisms. In the previous conditions, cloud computing framework allows enterprises with less investment to ecommerce business applications (B2B & B2C) [9].

Cloud computing has proved to very beneficial for the ecommerce platforms due to the following advantages [10]:

A. Cost-effective:

Cloud-based initiatives tend to be more cost-effective for ecommerce organizations since all set up and maintenance activities because of their business are undertaken by way of the cloud vendors. For the reason, that cloud vendor specifically dedicates these activities for assorted clients, the all inclusive costs is shipped among these and therefore might be more cost-effective with the retailers [7].

B. Speed of Operations:

Installation and execution of e-commerce platform using cloud platforms is fast because of the fact that IT infrastructure needed for hosting the approval has already been installed. What's more, it hastens the execution use of various modules of the approval and it is more capable.

C. Scalability:

In Cloud computing platforms, the resource utilization and requirement for virtually every platform can be simply of scaled up or down. This will aid in handling the expenditure of hosting tweaking the working platform for your retailers. Also, scalability contributes in optimizing the burden time of the necessary paperwork while using the traffic. Hence, cloud service is again less expensive for your retailer.

D. Security:

The hearts for information loss or network intrusion has become checked effectively with the roll-out of various standards drafted by various organizations like ISO for cloud vendors. The vendors who go through the standards are simply just permitted to provide such service along with improved customer is important the reasoning behind contributes to picking solely those vendors which have been certified by such organizations [7].

6.1 Major Benefits of Cloud-Based E-Commerce Applications

- It allow organizations to respond quickly to available opportunities and challenges
- It enables IT and business players to evaluate new engagements without any huge funds
- Consumerization of the online client experience needs closer scrutiny of solution offerings
- IT leaders must understand the pros and cons of cloud-based ownership models in order to select the right solution for their needs

VII. DATA MINING INTEGRATION IN CLOUD

In Microsoft, a Microsoft suite of cloud-based services integrates a latest technical preview of Data Mining in the Cloud known as DMCloud. DMCloud gives permissions to you to manage some basic data mining tasks leveraging a cloud-based Analysis Services connections.

DMCloud is most valuable capability for IWs that would like to begin as long as SQL Server Data Mining without the added load of needing a technology professional to first installs Analysis Services. Furthermore, IWs can use the DMCloud services no matter where they may actually be located as long as they have Internet connections! The data mining functionality you can maintain with DMCloud are the same Table Analysis Tools found in the modern Excel Data Mining add-in. Thus, these data mining tasks include [8]:

- Detect Categories

- Shopping Basket Analysis
- Scenario Analysis
- Fill From Example
- Highlight Exceptions
- Forecast
- Analyze Key Influencers
- Prediction Calculator

VIII. CONCLUSION

Data mining is used in multiple applications such as Student management systems, Health care, Science, mathematics and in various e-commerce and online shopping websites. Cloud Computing proves the modern trend in Internet services that based on clouds of servers to handle tasks and functionalities. The future of data mining with cloud comes in predictive analytic. Data mining in cloud computing is the mechanism of distributing structured information from unstructured or semi-structured web data resources. This is very helpful for e-commerce activities. Data mining in Cloud Computing permits organizations to consolidate the management of software and storage of related data, with security of sufficient, sincere and secure services for their customers. Here we inspect how the data mining tools like PAS, SAS, and IaaS are used in cloud computing to distribute the information. Users use this feature to develop information listings, and to get information about different topics by searching in forums. Company's use this service to search what kind of information is floating across in the World Wide Web for their products or services and take actions based on the data granted. Furthermore, the information retrieval practical model through the multi-agent system with data mining in a cloud computing framework has been create and proposed. However, it's suggested that users should arrange that the request made to the IaaS is within the scope of integrated data warehouse and is very simple and clear. Henceforth, to providing the work for the multi-agent system easier through application of the data mining algorithms, fetch meaningful information from the data warehouses. Cloud computing admit the users to retrieve meaningful information from virtually integrated data warehouse that decreases the expenditures of infrastructure, framework and storage.

ACKNOWLEDGEMENT

We would like to gratefully and sincerely thank to The Dean of our College, and Chairman of our Departments for their guidance, understanding, patience, and most importantly, their friendly nature during this research paper writing at College of computers and Information Technology, Taif University. We would also like to thank all of the members of the research group and friends who have always supported us for giving the opportunity to write this research paper.

REFERENCES

- [1] Alex Berson, Stephen J.Smith, Berson, Kurt Thearling. Building Data Mining Applications for CRM [Paperback].
- [2] Roger Jennings . Cloud Computing with the Windows Azure Platform.
- [3] Dinkar Sitaram, Geetha Manjunath. Moving To the Cloud: Developing Apps in the New World of Cloud Computing.
- [4] Todd Arias. The Cloud Computing Handbook - Everything You Need to Know about Cloud Computing.
- [5] <http://searchsqlserver.techtarget.com/definition/data-mining>
- [6] Venkatadri.M, Dr. Lokanatha C. Reddy. A Review on Data mining from Past to the Future
- [7] <http://dssresources.com/books/contents/berry97.html>
- [8] <http://www.marketingprofs.com/articles/2010/3567/the-nine-most-common-data-mining-techniques-used-in-predictive-analytics>
- [9] Ahmed Abou Elfetouh Saleh. A proposed framework based on cloud computing for enhancing e-commerce applications. *International Journal of Computer Applications* 59(5):21-25, December 2012.
- [10] Xiaofeng Wang. Research on e-commerce development model in the cloud computing environment. *International Conference on System*

Science and Engineering (ICSSE), 2012.

[11] Mohd Muntjir: Novice Evaluation and Comparative Survey on Database Management System, Data Warehousing and Data Mining. *International Journal of Computer Applications* 136(10):39-45, February 2016. Published by Foundation of Computer Science (FCS), NY, USA

[12] Mohd Muntjir and Sultan Aljahdali: DBMS Integration With Cloud Computing, *European Journal of Computer Science and Information Technology*, 1(1):23-29, June 2013, Published by European Centre for Research Training and Development UK

[13] Ahmad Tasnim Siddiqui and Amjath Fareeth Basha. E-Business Strategies To Cut Back Cost of Business Enterprise, 3rd IEEE International Advance Computing Conference (IACC) 2013, 520-524, June 2013, Published by IEEE.

Radial Basis Polynomial Kernel (RBPk): A Generalized Kernel for Support Vector Machine

*Ms. Hetal Bhavsar¹, Dr. Amit Ganatra²

¹Assistant Professor, Department of Computer Science & Engineering,
The M. S. University of Baroda, Vadodara, India.

²H. O. D., Computer Engineering Department,
Charotar University of Science & Technology, Changa, Dist. Anand, India.

Abstract:

Support Vector Machine (SVM) is a novel machine learning method, based on the statistical learning theory and VC (Vapnik-Chervonenkis) dimension concept. It has been successfully applied to numerous classification and pattern recognition problems. Generally, SVM uses the kernel functions, when data is non-linearly separable. The kernel functions map the data from input space to higher dimensional feature space so the data becomes linearly separable. In this, deciding the appropriate kernel function for a given application is the crucial issue. This research proposes a new kernel function named “Radial Basis Polynomial Kernel (RBPk)” which combines the characteristics of the two kernel functions: the Radial Basis Function (RBF) kernel and the Polynomial kernel and proves to be better kernel function in comparison of the two when applied individually. The paper proves and makes sure that RBPk confirms the characteristics of a kernel. It also evaluates the performance of the RBPk using Sequential Minimal Optimization (SMO), one of the well known implementation of SVM, against the existing kernels. The simulation uses various classification validation methods viz. holdout, training vs. training, cross-validation and random sampling methods with different datasets from distinct domains to prove the usefulness of RBPk. Finally, it concludes that the use of RBPk results into better predictability and generalization capability of SVM and RBPk can become an alternative generalized kernel.

Keywords: Support vector machine; kernel function; sequential minimal optimization; feature space; polynomial kernel; and Radial Basis function

1. Introduction

Support Vector Machine (SVM) is a supervised machine learning method, based on the statistical learning theory and VC dimension concept. It is based on structural risk minimization principle which minimizes an upper bound on the expected risk, as opposed to empirical risk minimization principle that minimizes the error on the training data. It uses a margin-based criterion that is attractive for many classification applications like Handwritten digit recognition, Object recognition, Speaker Identification, Face detection in images, text categorization, Image classification, Bio-sequence analysis etc. [8], [13], [28], [29]. It also follows Generalization and regularization theory, which gives the principle way to choose a hypothesis [6], [7]. Training a Support Vector Machine comprises of solving large quadratic programming (QP) problems in order to find the optimal separating hyper-plane, which requires $O(m^2)$ space complexity and $O(m^3)$ time complexity, where m is the number of training samples [8], [15].

Traditionally, SVM was used for linearly separable datasets to find the optimal separating hyper-plane (dark line in Fig. 1) from the large number of separating hyper-plane, that optimally separate the data into two areas as shown in figure.

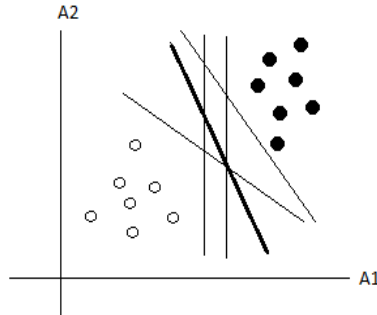


Figure 1: Number of Separating Hyper-plane

However, in the real world, all the datasets for instance, Iris [1], x-or [13] are not linearly separable, always. For such datasets, SVMs are extended with the Kernel functions. Kernel functions map the data from input space to higher dimensional feature space. It is also proved in Cover's theorem that any dataset becomes arbitrarily separable as the data dimension grows [21]. Thus the mapping of nonlinear separable datasets into the higher dimensional feature space makes classification problem linear [8]. Figure 2 shows mapping of non-linear datasets to higher dimensional feature space. SVM, then, finds the hyper-plane of maximal margin in the new feature space.

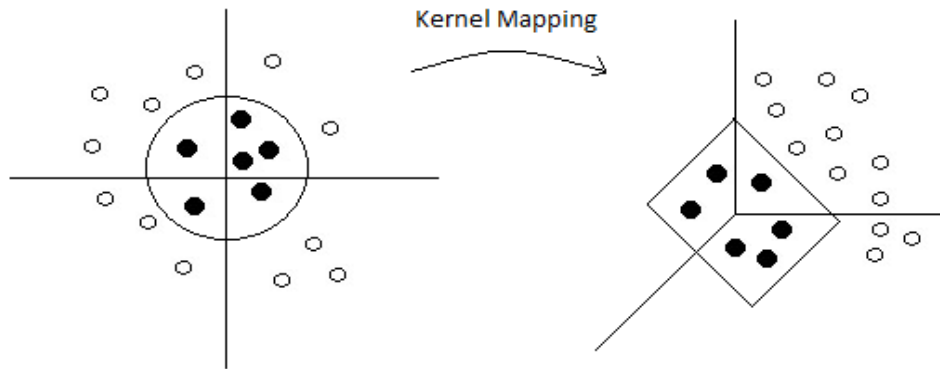


Figure 2: Mapping non-linear data to higher dimensional feature space

Let the Data set D be given as $\{x_i, y_i\}$, where $x_i, i=1,2,\dots,l$ (i.e. no. of samples or training tuples) and $y_i \in \{-1,1\}$ and $x_i \in \mathbb{R}^d$. If the training data is non-linearly separable, then the QP problem for training an SVM in dual space, with kernel function to find an optimal hyper-plane is

$$W(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j K(\vec{x}_i, \vec{x}_j) \quad (1)$$

Subject to: $\sum_{i=1}^l \alpha_i y_i = 0$ and $C \geq \alpha_i \geq 0$

And the classification function of SVM is defined as:

$$f(x) = \text{sgn}(\sum_{i=1}^{Ns} \alpha_i y_i K(\vec{x}_i, \vec{x}) + b) \quad (2)$$

where C is the regularization parameter, $K(\vec{x}_i, \vec{x}_j)$ is the kernel function, which measures the similarity or distance between the two vectors, and the variables α_i are Lagrange multipliers.

Since the feature space of every kernel function is different, the representation of data in the feature space is also different. Deciding which kernel is the most suitable for a given application is an important issue and difficult as well. Along with this, for every kernel function, tuning of the parameters is required for better efficiency of the classification problem and hence, determining the optimal combination of kernel function and parameters is crucial for any SVM classification problem.

Linear kernel, Polynomial kernel, Radial Basis Function (RBF) and Sigmoid kernel are the common and well known kernel functions. This research analyzes the key characteristics of two kernel functions, RBF kernel and Polynomial kernel, and proposes a new kernel function, which named as “Radial Basis Polynomial Kernel” (RBPk). It combines the advantages of the two, resulting into better learning and better predictability. Experiments done on the real-world datasets of different domains with the four validation methods Holdout, Training vs. Training, Cross Validation and Random Sampling shows the conformance of RBPk. The better performance of RBPk on different datasets proves that it can be served as a generalized kernel.

To make SVM more space and time efficient, many algorithms and implementation techniques have been developed to train SVM for massive datasets. Decomposition techniques speed up the SVM training by dividing the original QP problem into smaller pieces, thereby reducing the size of each QP problem. Chunking algorithm, Osuna’s decomposition algorithm are well known decomposition algorithms [14]. Since these techniques require many passes over the data set, they need a longer training time to reach a reasonable level of convergence.

Therefore, for performance evaluation, SMO implementation of SVM is used. SMO is a special case of decomposition methods where in each sub problem two coefficients are optimized per iteration which is solved analytically. The advantages of SMO are: It is simple, easy to implement, generally faster, and has better scaling properties for difficult SVM problems than the standard SVM training algorithm [11], [14]. It maintains kernel matrix of size which equal to total number of samples in dataset and thus scales between linear and cubic in the sample set size.

To find an optimal point of equation (1), SMO algorithm uses the Karush-Kuhn-Tucker (KKT) conditions. The KKT conditions are necessary and sufficient conditions for an optimal point of a positive definite QP problem. The QP problem is solved when, for all i , the following KKT conditions are satisfied:

$$\left. \begin{aligned} \alpha_i &= 0 \Leftrightarrow y_i u_i \geq 1 \\ 0 < \alpha_i < C &\Leftrightarrow y_i u_i = 1 \\ \alpha_i &= C \Leftrightarrow y_i u_i \leq 1 \end{aligned} \right\} \quad (3)$$

where u_i is the output of the SVM for i^{th} training sample. The KKT conditions can be evaluated on one example at a time, which forms the basis for SMO algorithm. When it is satisfied by every multiplier, the algorithm terminates. The KKT conditions are verified to within ε , which typically range from 10^{-2} to 10^{-3} .

The organization of the paper is as follows. Section 2 gives an in depth study of the related work by the data mining community. The kernel function, RBF kernel and Polynomial kernel are explained in section 3. The section 4 describes the proposed RBPk function. The experimental results are discussed in section 5. Finally the paper concludes in section 6.

2. Related Work

Many researchers have already worked to propose a novel kernel function for improvement of SVM performance. The clinical kernel function which takes into account the type and range of each variable is proposed in [2]. This requires the specification of each type of variable, as well as the minimal and maximal possible value for continuous and ordinal variables based on the training data or on a priori knowledge. A new method of modifying a kernel to improve the performance of a SVM classifier which is based on information-geometric consideration of the structure of the Riemannian geometry induced by the kernel is proposed in [26]. The basic idea behind this kernel is to increase the separability between classes. A new kernel by using convex combination of good characteristics of polynomial and RBF kernels is proposed in [10]. To guarantee that the mixed kernel is admissible, optimal minimal coefficient has to be determined. The advantages of linear and Gaussian RBF kernel function are combined to propose a new kernel function in [12]. This results into the better capability of generalization and prediction, but the method they used to choose the best set of parameters (C , σ , λ) is time consuming, requiring $O(N^3)$ time complexity where N is the number of training samples. The compound kernel taking polynomial kernel, the RBF and the Fourier kernel is given in [30]. The use of this compound kernel posed a better learning as well as better extrapolation capability as compared to single kernel function. Still, it does not solve the problem of selection

of kernel function for improving the performance of SVM. The Minimax probability machine (MPM) whose performance depends on its kernel function is evaluated by replacing Euclidean distance in the Gaussian kernel with a more generalized Minkovsky's distance, which result into better prediction accuracy than the Euclidean distance [31]. In [25], the authors propose dynamic SVM by distributing kernel function showing that recognition question of a target feature is determined by a few samples in the local space taking it as the centre and the influence of other samples can be neglected. [23] Shows that there may be the risk of losing information while multiple kernel learning methods try to average out the kernel matrices in one way or another. In order to avoid learning any weight and suit for more kernels, the new kernel matrix is proposed which composed of the original, different kernel matrices, by constructing a larger matrix in which the original ones are still present. The compositional kernel matrix is s times larger than the base kernels. A new mechanism to optimize the parameters of combined kernel function by using large margin learning theory and a genetic algorithm, which aims to search the optimal parameters for the combined kernel function is proposed in [19]. However, the training speed is slow when the dataset becomes large. The influence of the model parameters of the SVMs using RBF and the scaling kernel function on the performance of SVM are studied by simulation in [32]. The penalty factor is mainly used to control the complexity of the model and the kernel parameter mainly influences the generalization of SVM. They showed that when the two types of parameters function jointly, the optimum in the parameter space can be obtained.

Though RBF kernels are widely used, the major problem is the selection of kernel and margin parameter. This can be done by time consuming grid search method. The generalized RBF kernels, called Mahalanobis kernel, is proposed in [24]. The covariance matrix for Mahalanobis kernel is calculated using the training data and the use the line search method to find optimum value of margin and kernel parameter instead of costly grid search algorithm. An improved multi-kernel Least Square SVM (LS-SVM) which combines the advantages of linear and Gaussian kernel is proposed in [20]. They showed that the proposed method gives better performance compared to existing LS-SVM method which uses single kernel function. To obtain the optimal free parameter, it uses Constrained Particle Swarm Optimization.

Along with the kernel functions, the values of parameters of kernel function's (like d in polynomial kernel, σ in RBF function and p in sigmoid kernel) and regularization parameter C has a great impact on complexity and generalization error of the classifier. Choosing the optimal values of these parameters is also very important along with the selection of kernel function [4], [12].

Many authors have suggested a variety of the ways for the kernel parameters selection. One of the ways is mentioned in [9] where grid search algorithms have been used to find the best combination of the SVM kernel and parameters. These algorithms are iterative and have been proven computationally costlier during the training phase and hence efficiency of the SVM classifier has been severely degraded. The evolutionary algorithm, in [3], is proposed to optimize the SVM parameters, including kernel type, kernel parameters and regularization parameter C , which is based on the genetic algorithm. It uses, repeatedly, the process of the crossover, mutation and selection to produce the optimal set of parameters. In this, the convergence speed depends on the crossover, mutation and selection functions. In [16], the method that avoids the iterative process of evaluating the performance for all the parameter combination is proposed. This approach selects the kernel parameter using the distance between two classes (DBTC) in the feature space. The optimal parameters are approximated with the sigmoid function and the computational complexity decreases significantly.

In [17] also, a method using the inter-cluster distances in the feature spaces, to choose the kernel parameters for training the SVM models, is proposed. Calculating the inter-cluster distance takes much lesser computation time than training the corresponding SVM classifiers; thus the proper kernel parameters can be chosen much faster. With properly chosen distance indexes, the proposed method performs stable with different sample sizes of the same problem. However, the penalty parameter is not incorporated into the proposed strategies in which the training time of SVM might be further minimized. A new feature weight learning method for SVM classification is introduced in [27]. The basic idea of this method is to tune the parameters of the Gaussian ARD kernel via optimization of kernel polarization. Each tuned parameter indicates the relative importance of the corresponding feature. Experimental results on some real data sets show that, this method leads to improvement in the classification accuracy and reduction in the number of support vectors, both.

However, the choice of the SVM kernel function and its parameter are still a relatively complex and difficult issues. This research analyzes the key characteristics of two very well known kernel functions: the RBF kernel and the

Polynomial kernel, and proposes a new kernel function combining the advantages of the two, which gives better learning and better prediction ability.

The Table 1 summarizes some of the well known kernel functions:

Table 1: List of some well-known kernel functions

Sr. No.	Year	Author	Kernel Used
1	1999[26]	S. Amari, S.Wu	$\tilde{K}(\mathbf{x}, \mathbf{x}') = c(\mathbf{x})c(\mathbf{x}')K(\mathbf{x}, \mathbf{x}')$ is called a conformal transformation of a kernel by factor $c(\mathbf{x})$.
2	2002[10]	G.F. Smits, E.M Jordaan	$K_{mix} = \rho K_{poly} + (1 - \rho)K_{rbf}$
3	2008[12]	H. Song, Z. Ding, C. Guo, Z. Li, and H. Xia	$K(x_i, x_j) = (1 - \lambda) \cdot (x_i \cdot x_j) + \lambda \cdot e^{-\left(\frac{\ x_i - x_j\ ^2}{\sigma^2}\right)}$
4	2008[19]	M. Lu, C. P. Chen, J. Huo, and X. Wang	$K_{combined} = (K_1)^{e_1} \circ \dots \circ (K_m)^{e_m}$ Where where K_i (1, 2,..., m) denotes the kernel function i, e is the exponent of i -th kernel function and \circ represents the operator between the two kernel functions, which can be addition and multiplication operators.
5	2009[2]	A. Daemen and B. De Moor	For continuous and ordinal variable $k_x(i, j) = \frac{(max - min) - x_i - x_j }{max - min}$ For Nominal variable $k_x(i, j) = \begin{cases} 1 & \text{if } x_i = x_j \\ 0 & \text{if } x_i \neq x_j \end{cases}$ Final kernel is average of all variables
6	2009[31]	X. Mu and Y. Zhou	$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\ \mathbf{x} - \mathbf{x}'\ _2^2}{\sigma^2}\right) = \exp\left(-D_{Euc}^2(\mathbf{x}, \mathbf{x}') / \sigma^2\right)$ The Euclidean distance has a natural generalization in form of the Minkovsky's distance $D_{Mm}(\mathbf{x}, \mathbf{x}', \alpha) = \ \mathbf{x} - \mathbf{x}'\ _M = \left(\sum_{i=1}^d x_i - x'_i ^\alpha\right)^{\frac{1}{\alpha}}$
7	2010[30]	W. An-na, Z. Yue, H. Yuntao, and L. Yun-lu	$K(x, x') = \rho_1((x \cdot x') + 1)^d + \rho_2 \exp\left(-\frac{\ x - x'\ ^2}{2\sigma^2}\right) + \rho_3 \tanh(v(x \cdot x') + c)$ parameters (ρ_1, ρ_2, ρ_3) are the proportion of the above
8	2010[25]	S. Guangzhi, D. Lianglong, H. Junchuan, and Z. Yanxia	$K(x_i, x_j) = \exp\left\{-\frac{ x_i - x_j ^2}{\sigma_i \sigma_j}\right\}$ Distributing kernel function, Where σ_i is measured by using the distance between the target feature and each training sample
9	2010[23]	R. Zhang and X. Duan	$K_p^{\frac{1}{2}}(x, x') K_{p'}^{\frac{1}{2}}(x, x')$
10	2011[32]	Y. Jin, J. Huang, and	

		J.Zhang	$K(x_i, x_j) = \prod_{m=1}^n \frac{\sin[P(x_i^m - x_j^m)]}{P(x_i^m - x_j^m)}$
11	2005 [24]	S. Abe	$H(x, x') = \exp[-(x - x')^T A (x - x')]$
12	2011 [20]	M. Tarhouni, K. Laabidi, S. Zidi and M. Ksouri-Lahmari	$K(x, x_i) = a_1(xx_i^T + cst) + a_2 \exp(-\frac{\ x - x_i\ ^2}{2\sigma^2})$

3. Polynomial and RBF Kernel FUNCTIONS

The Kernel functions are used in Support Vector Machines to map the nonlinear inseparable data into a higher dimensional feature space where the computational power of the linear learning machine is increased [15], [18].

Let x_i be a vectors in the d-dimensional input space as defined above and Φ be a nonlinear mapping function from input space to (possibly infinite dimension) Euclidean space \mathcal{H} (feature space) denoted by:

$$\Phi: \mathbb{R}^d \rightarrow \mathcal{H} \text{ Where } x_i \in \mathbb{R}^d \text{ and } \Phi(x_i) \in \mathcal{H} \quad (4)$$

\mathcal{H} can also be referred to as Hilbert space, which can be thinking as a generalization of Euclidean space. With this transformation of data in high dimensional feature space, the dual problem of SVM can be written as follows,

$$w(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \Phi(\vec{x}_i) \cdot \Phi(\vec{x}_j) \quad (5)$$

and the decision function of SVM can be defined as,

$$f(x) = \text{sgn}(\sum_{i=1}^{N_s} \alpha_i y_i \Phi(\vec{x}_i) \cdot \Phi(\vec{x}_j) + b) \quad (6)$$

Here, the feature mapping function is always appearing as dot products, i.e. $\Phi(\vec{x}_i) \cdot \Phi(\vec{x}_j)$, as shown in equation (5) and (6).

Depending on the chosen value of Φ , \mathcal{H} might be high or even infinite dimensional, so computing this inner product in transformed feature space is quite complex, costly and suffer from curse of dimensionality. However, we do not require the dot product of $\Phi(\vec{x}_i)$ and $\Phi(\vec{x}_j)$ in high dimensional feature space, if we find a kernel function $K(\vec{x}_i, \vec{x}_j)$ such that,

$$K(\vec{x}_i, \vec{x}_j) = \Phi(\vec{x}_i) \cdot \Phi(\vec{x}_j) \quad (7)$$

It measures the similarity or distance between the two vectors. Finding such kernel function, we can calculate the dot product of $(\Phi(\vec{x}_i), \Phi(\vec{x}_j))$ without explicitly applying function Φ to input vector.

Any algorithm for vectored data which is expressed in terms of dot products between vectors can be performed implicitly in the feature space associated with any kernel by replacing each dot product by a kernel evaluation [21]. Hence, replacing $\Phi(\vec{x}_i) \cdot \Phi(\vec{x}_j)$ by kernel function, the dual problem of SVM (equation (5)) is defined as equation (1) and the classification function of SVM (equation (6)) is defined as equation (2).

The kernel function can transform the dot product operations in high dimension space into the kernel function operations in input space as long as it satisfies the Mercer condition [4], [6], [22]; thereby it avoids the problem of computing directly in high dimension space and solves the dimension adversity.

The performance of SVM largely depends on the kernel function. Every kernel function has its own advantages and disadvantages. Various possibilities of kernels exist and it is difficult to explain their individual characteristics. A single kernel function may not have a good learning as well as generalization capability. As a solution, the good characteristics of two or more kernels should be combined.

Generally, there are two types of Kernel functions of support vector machine: local kernel function and global kernel function. In global kernel function samples far away from each other has impact on the value of kernel function; however in local kernel function only samples closed to each other has impact on the value of kernel function. This research has considered, one of the global kernel function i.e. Polynomial kernel and one of the local kernel function i.e. the RBF kernel, and try to combine their advantages which result into a kernel function in which all the samples (far or near) has impact on the value of kernel function.

a. Polynomial kernel

The polynomial kernel function is defined as

$$k(\vec{x}_i, \vec{x}_j) = (\vec{x}_i \cdot \vec{x}_j + 1)^d \quad (8)$$

Where d is the degree of the kernel. In Figure 3 the global effect of the Polynomial kernel of various degrees is shown over the data space [-1, 1] with test input 0.2, which shows that every data point from the data space has an influence on the kernel value of the test point, irrespective of its actual distance from test point.

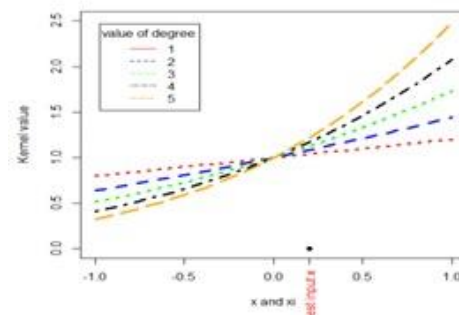


Figure 3: A global polynomial kernel function with different values of d

The polynomial kernel has a good generalization ability, which can affect the value of global kernel. The poor leaning ability is the disadvantage of polynomial kernel.

b. RBF kernel

The RBF kernel function is the most widely used kernel function because of its good learning ability among all the single kernel functions.

$$k(\vec{x}_i, \vec{x}_j) = e^{-\|\vec{x}_i - \vec{x}_j\|^2 / 2\sigma^2} \quad (9)$$

where, $\sigma^2 = \text{mean}\|\vec{x}_i - \vec{x}_j\|^2$

The RBF can be well adapted under many conditions, low-dimension, high-dimension, small sample, large sample, etc. RBF has the advantage of having fewer parameters. A large number of numerical experiments proved that the learning ability of the RBF is inversely proportional to the parameter σ . σ determines the area of influence over the data space. Figure 4 shows the local effect of RBF kernel for a chosen test input 0.2 over the data space [-1, 1], for different values of the width σ . A larger value of σ will give a smoother decision surface and more regular decision boundary. This is because an RBF with large σ allow a support vector to have a strong influence over a large area. If σ is very small, we can see in fig. 1 that only samples whose distances are close to σ can be affected. Since, it affect on the data points in the neighbourhood of the test point, it can be call as local kernel.

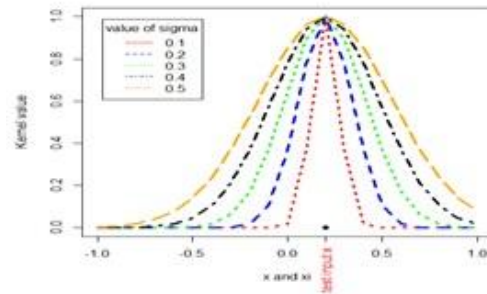


Figure 4: A local RBF kernel function with different value of σ

4. The RBPK FUNCTION

For a SVM classifier, choosing a specific kernel function means choosing a way of mapping to project the input space into a feature space. A learning model, which is judged by its learning ability and prediction ability, was built up by choosing a specific kernel function. Thus, to build up a model which has good learning as well as good prediction ability, this research has combined the advantages of both, local RBF kernel function and global Polynomial kernel function.

A novel kernel function called Radial Basis Polynomial Kernel (RBPK) is now defined as:

$$k(x_i, x_j) = \exp\left(\frac{(x_i \cdot x_j + c)^d}{\sigma^2}\right) \quad (10)$$

Where $c > 0$ and $d > 0$.

The RBPK takes advantage of good prediction ability from polynomial kernel and good learning ability from RBF kernel function.

The Mercer's theorem provides the necessary and sufficient condition for a valid kernel function. It says that a kernel function is a permissible kernel if the corresponding kernel matrix is symmetric and positive semi-definite (PSD) [4], [7], [22]. It can be determined that a kernel matrix is PSD by determining its spectrum of eigenvalues. Note that a symmetric matrix is positive definite if and only if all its Eigen values are non-negative. According to this, to be a permissible kernel, the RBPK must satisfy the Mercer's theorem.

To prove the validity of RBPK, equation (10) can be expanded as (according to [22]),

$$= \sum_{i=0}^{\infty} \frac{1}{\sigma^{2i} i!} (x_i \cdot x_j + c)^{d \cdot i} \quad (11)$$

$$= 1 + \sum_{i=0}^{\infty} \frac{1}{\sigma^{2i} i!} (x_i \cdot x_j + c)^{d \cdot i} \quad (12)$$

Using proposition 3 and 4 of Theorem 2.20, Mercer's conditions are proved true for RBPK and hence the proposed RBPK is a valid kernel [22].

5. Experimental Results an discussion

The validity of RBPK has been tested and evaluated using several datasets against the existing kernel functions like the linear, the polynomial and the RBF. Datasets used for experiments are from different domains of varying

characteristics and taken from UCI [1] and LIBSVM dataset repository [5]. Details of these datasets are shown in Table 2.

Table 2: Datasets used for RBPK

Dataset	No. of Classes	No. of Features	No. of Training Instances	No. of Testing Instances	Domain
Iris	3	4	150	-	Plants
Heart	2	13	270	-	Medical
Glass	7	10	214	-	Physical
Adult(a1a)	2	123	1605	30956	Personal
DNA	3	180	2000	1186	Medical
Letter	26	16	15000	5000	Pattern
USPS	10	256	2007	7291	Pattern
web8	2	300	45546	13699	Web
Coverttype	2	54	581012	-	Life

a. Experimental Setup and Preliminaries

The experiments have been conducted using the LIBSVM implementation of SMO with linear, polynomial, RBF and proposed RBPK functions. The classification accuracy of this four kernel functions are measure with four methods: 1) Holdout 2) Training vs. training 3) Cross-validation and 4) Random sampling. With holdout method, the dataset is split into two sets: 1) training set, which is used to train the classifier and 2) testing set, which is used to estimate the error rate of the trained classifier. Generally, two-thirds of data are selected for training and one-third of data is selected for testing, and hence, 60% of samples are used for training and 40% of samples are used for testing for Iris, Heart and Glass dataset. For other datasets training and testing datasets are already available as separate files. In training vs. training method, the same set of samples is used for training as well as for testing of the classifier. In k-fold cross validation, data is split into k disjoint subsets (folds). Training and testing is performed k times. For each of this k experiment, k-1 folds are used for training and remaining one is used for testing. The error rate of the classifier is the average of the error rates of k experiments. For large datasets like Adult, USPS, Web8 and Coverttype, number of folds are set to 5 and for other datasets; the number of folds are set to 10. With the random sampling method, random division of datasets into training and testing samples are done

To make the choice of the free parameters, we considered the following combinations of values:

Complexity or Regularization Parameter C	1
Kernel type	Linear Polynomial RBF RBPK
Polynomial – Parameter (degree – d)	3 5
RBF – Parameter (gamma – γ)	0.01 0.05 0.08 0.1
RBPK – Parameter (γ and d)	Same as above

The parameter values are considered from the intervals where the SVM method showed a regularly nearly optimal behavior in terms of testing classification accuracy and number of SVs. The result tables include only the best result obtained for different kernels after the parameter tuning. The measures like accuracy, training time, testing time, number of correctly classified instances (CCI), precision, True Positive Rate (TPR) and False Positive Rate (FPR) are used to compare the performance of the Linear, the Polynomial, the RBF and the RBPK.

The simulation results taken by running SMO algorithm using LIBSVM framework in eclipse on Intel core i5-2430M CPU@ 2.4GHz with 4GB of RAM Machine.

5.1 Result for IRIS dataset

Table 3: Experimental results on IRIS dataset

Methods	Kernel Function	Parameters	Accuracy	Training Time (Sec)	Testing Time (Sec)	Correctly Classified Instances (CCI)	Number of SVs	Precision	TPR	FPR
Holdout (90/60)	Linear		93.33	0.016	0	56	34	0.924	0.924	0.034
	Polynomial	d=3	68.33	0.016	0.016	41	81	0.829	0.677	0.157
	RBF	$\gamma=0.08$	93.33	0	0.016	56	60	0.927	0.925	0.033
	RBPK	d=5, $\gamma=0.08$	96.67	0.015	0.016	58	24	0.958	0.958	0.016
Training vs. Training (150/150)	Linear		97.33	0.008	0.009	146	42	0.964	0.963	0.013
	Polynomial	d=3	75.33	0.01	0.009	113	122	0.85	0.746	0.122
	RBF	$\gamma=0.08$	96.67	0.008	0.01	145	85	0.958	0.957	0.016
	RBPK	d=5, $\gamma=0.08$	98.67	0.011	0.013	148	45	0.967	0.976	0.006
Cross Validation (k=10) (150)	Linear		97.33	0.02	-	146	40	0.974	0.973	0.012
	Polynomial	d=3	73.33	0.026	-	110	113	0.944	0.734	0.02
	RBF	$\gamma=0.08$	96.66	0.027	-	145	79	0.958	0.957	0.015
	RBPK	d=5, $\gamma=0.08$	98	0.022	-	147	31	0.98	0.98	0.01
Random Sampling (90/60)	Linear		96.67	0	0.016	58	30	0.967	0.967	0.016
	Polynomial	d=3	66.67	0.016	0	40	80	0.506	0.67	0.156
	RBF	$\gamma=0.05$	96.67	0.016	0.015	58	69	0.967	0.967	0.016
	RBPK	Default	98.33	0.016	0	59	22	0.984	0.983	0.008

As seen from Table 3, For Iris dataset, the accuracy of RBPK is highest compared to linear, polynomial and RBF kernel function for all the four methods. The accuracy of RBPK is increased by 0.7% to 3.5%. Also, the number of support vectors for RBPK is less compared to other kernel functions, which reduces the testing time complexity of SVM with RBPK. The Precision and TPR is also highest for RBPK compared to other kernel functions. Figure 5 shows comparison of accuracy and no. of SVs for IRIS dataset.

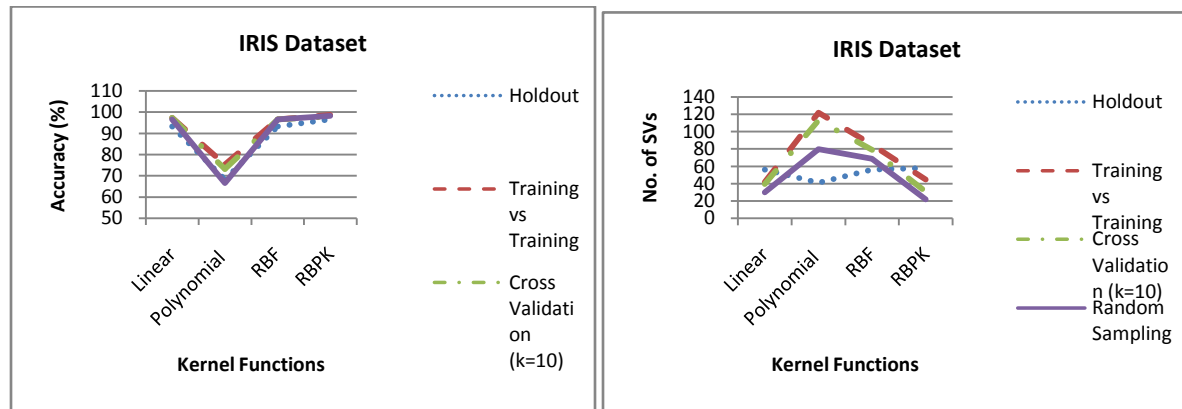


Figure 5: Accuracy and No. of SVs for IRIS dataset

5.2 Result for Heart dataset

Table 4: Experimental results on Heart Dataset

Methods	Kernel Function	Parameters	Accuracy	Training Time (Sec)	Testing Time (Sec)	Correctly Classified Instances (CCI)	Number Of SVs	Precision	TPR	FPR
Holdout	Linear		87.04	0.016	0	94	71	0.874	0.87	0.124

(162/108)	Polynomial	d=3	85.19	0	0	92	127	0.853	0.852	0.148
	RBF	$\gamma=0.01$	89.81	0	0	97	112	0.898	0.899	0.107
	RBPK	d=3, $\gamma=0.01$	89.81	0	0	97	130	0.898	0.899	0.107
Training vs. Training (270/270)	Linear		84.81	0.016	0.015	229	101	0.848	0.848	0.158
	Polynomial	d=3	85.93	0.014	0.015	232	177	0.859	0.86	0.147
	RBF	$\gamma=0.1$	87.03	0.015	0.016	235	133	0.872	0.871	0.143
	RBPK	d=5, $\gamma=0.05$	100	0.026	0.015	270	155	1	1	0
Cross Validation (k=10) (270)	Linear		84.07	0.067	-	227	91	0.842	0.841	0.161
	Polynomial	d=3	82.96	0.062	-	224	158	0.832	0.83	0.172
	RBF	$\gamma=0.01$	83.7	0.078	-	226	144	0.843	0.837	0.158
	RBPK	d=3, $\gamma=0.01$	84.44	0.078	-	228	108	0.848	0.845	0.155
Random Sampling (162/108)	Linear		85.19	0	0	92	67	0.852	0.852	0.852
	Polynomial	d=5	85.19	0	0	92	154	0.856	0.852	0.144
	RBF	$\gamma=0.1$	87.96	0	0	95	90	0.88	0.88	0.125
	RBPK	d=1, $\gamma=0.05$	86.11	0.015	0	93	81	0.86	0.861	0.152

For Heart dataset, Table 4 shows that, the accuracy of RBPK is increased by 0% to 13% from the highest accuracy of RBF kernel for holdout, training vs. training and cross-validation method. For Holdout method, the accuracy of RBPK is same as the highest accuracy given RBF kernel, while for training vs. training method, the accuracy of RBPK is 100%, which is 13% more than the highest accuracy of RBF kernel. The number of support vectors for RBPK is less compared to polynomial and RBF kernel functions, but more compared to linear kernel function. Figure 6 shows comparison of accuracy and no. of SVs for Heart dataset.

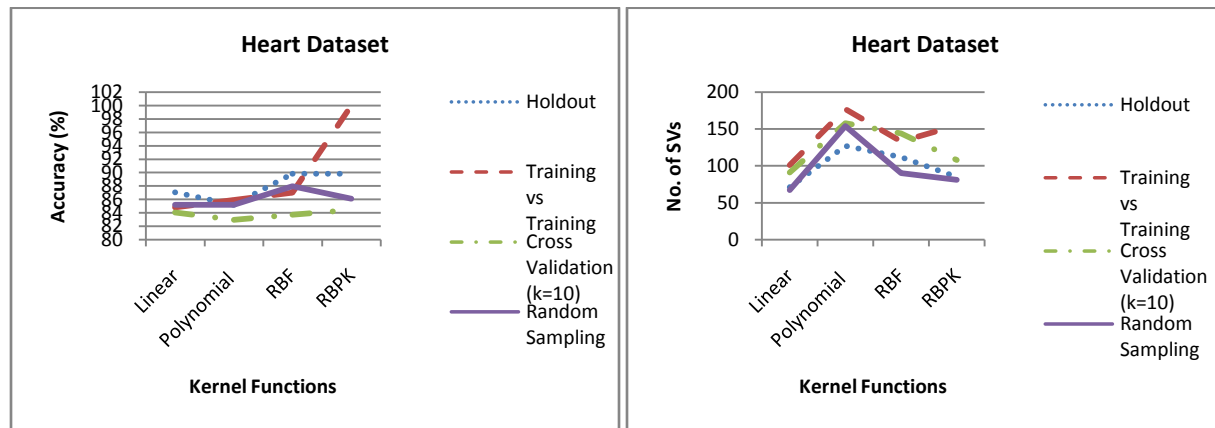


Figure 6: Accuracy and No. of SVs comparison for Heart Dataset

5.3 Result for Glass Dataset

The Table 5 gives an inference that, similar to the iris and heart datasets, for Glass dataset also, the accuracy of RBPK is highest compared to other kernel functions. For holdout, cross validation and random sampling the accuracy of RBPK increased from highest accuracy of linear kernel by 6% to 9%, while with training vs. training method, the accuracy of RBPK is 25% more than the highest accuracy of linear kernel. The number of support vectors for RBPK is least compared to all the other kernel function which result into the reduction of classification time. Figure 7 shows comparison of accuracy and no. of SVs for Glass dataset.

Table 5: Experimental Results on Glass Dataset

Methods	Kernel Function	Parameters	Accuracy	Training Time (Sec)	Testing Time (Sec)	Correctly Classified Instances (CCI)	Number of SVs	Precision	TPR	FPR
Holdout (128/86)	Linear		60.47	0.015	0	52	115	0.551	0.61	0.2
	Polynomial	d=3	38.37	0	0	33	128	0.226	0.385	0.327
	RBF	$\gamma=0.08$	50	0.016	0.016	43	123	0.451	0.504	0.249
	RBPK	d=5, $\gamma=0.1$	69.77	0.016	0	60	89	0.702	0.705	0.111
Training vs. Training (214/214)	Linear		66.82	0.032	0.016	143	184	0.631	0.678	0.163
	Polynomial	d=3	48.6	0.015	0.015	104	212	0.522	0.494	0.281
	RBF	$\gamma=0.1$	61.21	0.015	0.015	131	198	0.599	0.623	0.206
	RBPK	d=5, $\gamma=0.1$	92.06	0.032	0	197	131	0.93	0.93	0.038
Cross Validation (k=10) (214)	Linear		64.48	0.078		138	163	0.717	0.651	0.142
	Polynomial	d=3	46.26	0.094		99	189	0.962	0.462	0.007
	RBF	$\gamma=0.1$	58.88	0.093		126	179	0.772	0.597	0.136
	RBPK	d=5, $\gamma=0.08$	70.56	0.094		151	150	0.768	0.705	0.107
Random Sampling (170/44)	Linear		45.45	0.015	0	20	144	0.429	0.45	0.289
	Polynomial	d=3	45.45	0.016	0	20	169	0.254	0.448	0.309
	RBF	$\gamma=0.08$	54.55	0.015	0	24	161	0.538	0.54	0.251
	RBPK	d=5, $\gamma=0.08$	61.36	0.015	0	27	114	0.632	0.615	0.181

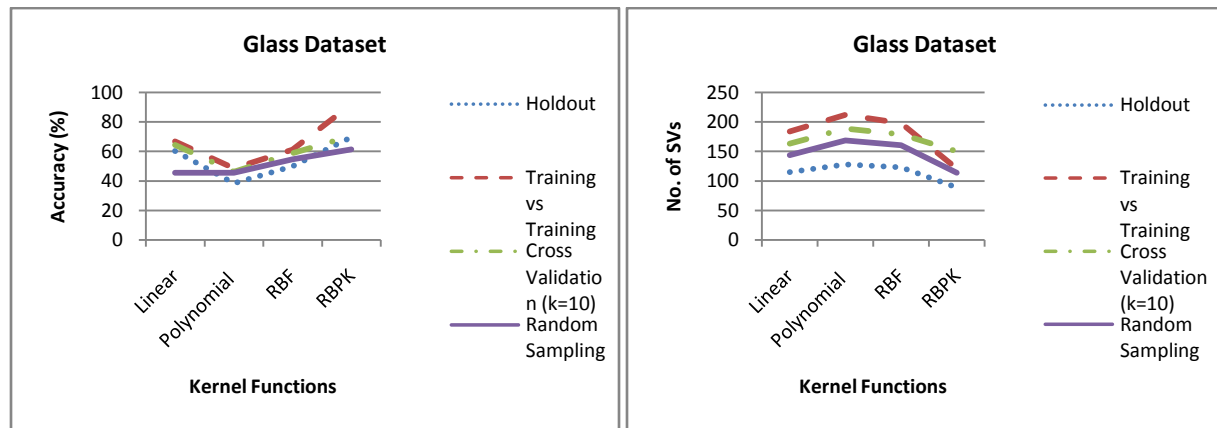


Figure 7: Accuracy and No. of SVs comparison for Glass Dataset

5.4 Result for Adult Dataset

Table 6, For Adult dataset, proves the accuracy of RBPK to be the highest accuracy as compared to the other kernel functions. For holdout, cross validation and random sampling the accuracy of RBPK increased from highest accuracy of RBF kernel by 0.06% to 0.58%, while with training vs. training method, the accuracy of RBPK is $\approx 9\%$ more than the highest accuracy of RBF kernel. Though the number of support vectors for RBPK is less than the polynomial and RBF kernel, the time taken to test the model is more for RBPK as the number of features for the dataset is large compared to iris, heart and glass datasets. Figure 8 shows comparison of accuracy and no. of SVs for Adult dataset.

Table 6: Experimental results for Adult Dataset

Methods	Kernel Function	Parameters	Accuracy	Training Time (Sec)	Testing Time (Sec)	Correctly Classified Instances (CCI)	Number of SVs	Precision	TPR	FPR
Holdout	Linear		83.82	0.244	2.568	25947	588	0.833	0.838	0.32

(1605/ 30956)	Polynomial	d=3	75.94	0.225	3.799	23510	804	0.577	0.76	0.76
	RBF	$\gamma=0.05$	84.23	0.24	4.144	26073	691	0.834	0.842	0.359
	RBPK	d=3, $\gamma=0.05$	84.31	0.271	4.413	26099	659	0.836	0.843	0.353
Training vs. Training (32561/ 30956)	Linear		84.99	154.23	55.775	27675	11519	0.843	0.85	0.323
	Polynomial	d=3	75.92	122.86	71.645	24720	15702	0.577	0.76	0.76
	RBF	$\gamma=0.1$	86.42	153.67	77.18	28138	11902	0.859	0.865	0.301
	RBPK	d=3, $\gamma=0.05$	95.6	1212.57	99.77	31128	13796	0.956	0.957	0.087
Cross Validation (k=5) (16280)	Linear		84.75	165.031		13798	4527	0.861	0.847	0.244
	Polynomial	d=3	75.91	167.55		12359	6287	1	0.759	0
	RBF	$\gamma=0.08$	84.82	173.14		13809	4879	0.866	0.848	0.239
	RBPK	d=5, $\gamma=0.01$	84.76	186.6		13799	4775	0.861	0.848	0.248
Random Sampling (10000/ 5000)	Linear		84.46	8.882	2.063	4223	3429	0.838	0.845	0.323
	Polynomial	d=3	75.92	9.103	3.3	3796	4834	0.577	0.76	0.76
	RBF	$\gamma=0.05$	84.66	8.823	3.316	4233	3597	0.84	0.847	0.332
	RBPK	d=3, $\gamma=0.01$	84.84	9.732	3.501	4242	3530	0.842	0.849	0.329

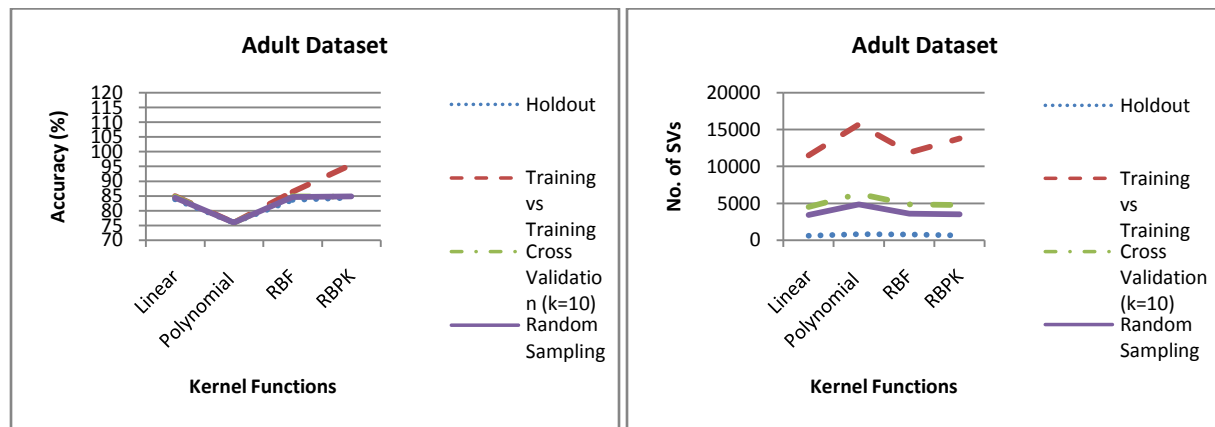


Figure 8: Accuracy and No. of SVs comparison for Adult Dataset

5.5 Result for DNA Dataset

The data values seen from Table 7, for DNA dataset infers that the accuracy of RBPK for holdout, cross validation and random sampling is increased about 0.5% from the highest accuracy of RBF kernel. Testing on the same dataset as the training, the accuracy of RBF and RBPK is similar which is 99.97%. Testing time of the model shows how it depends on the number of SVs. The FPR is lowest for RBPK compared to all other kernel function in all the four methods indicating that, very less number of samples is classified falsely. Figure 9 shows comparison of accuracy and no. of SVs for DNA dataset.

Table 7: Experimental Results for DNA dataset

Methods	Kernel Function	Parameters	Accuracy	Training Time (Sec)	Testing Time (Sec)	Correctly Classified Instances (CCI)	Number of SVs	Precision	TPR	FPR
Holdout (2000/ 1186)	Linear		93.08	0.749	0.312	1104	396	0.94	0.94	0.048
	Polynomial	d=3	50.84	2.575	1.263	603	1734	0.259	0.51	0.51
	RBF	$\gamma=0.01$	94.86	1.374	0.811	1125	1026	0.958	0.959	0.027
	RBPK	d=3, $\gamma=0.01$	95.36	2.06	1.03	1131	1274	0.963	0.963	0.026
Training vs.	Linear		99.34	1.618	1.004	3165	493	0.993	0.993	0.005
	Polynomial	d=3	51.91	6.93	5.389	1654	2733	0.27	0.52	0.52

Training (3186/)	RBF	$\gamma=0.1$	99.97	9.132	7.051	3185	3071	0.999	1	0.001
	RBPK	$d=3, \gamma=0.01$	99.97	4.078	3.985	3185	1807	0.999	1	0.001
Cross Validation (k=10) (3186)	Linear		93.47	14.48		2978	473	0.936	0.934	0.041
	Polynomial	$d=3$	51.91	65.42		1654	2463	1	0.519	0
	RBF	$\gamma=0.05$	96.12	73.634		3062	2185	0.961	0.961	0.023
	RBPK	$d=3, \gamma=0.01$	96.2	40.7		3065	1689	0.962	0.963	0.026
Random Sampling (2550/ 736)	Linear		92.8	0.891	0.219	683	416	0.928	0.928	0.045
	Polynomial	$d=3$	51.9	4.011	0.984	382	2117	0.27	0.52	0.52
	RBF	$\gamma=0.01$	95.11	1.848	0.594	700	1189	0.952	0.952	0.024
	RBPK	$d=3, \gamma=0.01$	95.52	2.564	0.735	703	1492	0.956	0.955	0.023

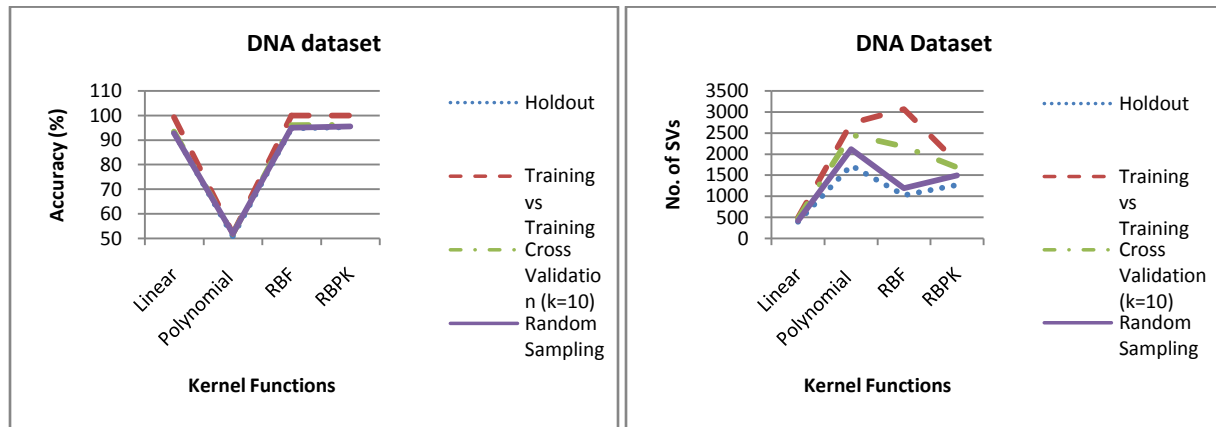


Figure 9: Accuracy and No. of SVs comparison for DNA Dataset

5.6 Result for Letter Dataset

The Letter dataset results, in Table 8, also prove RBPK as an accurate kernel with an increased accuracy of about $\approx 10\%$ to $\approx 12\%$ from the highest accuracy of RBF and linear kernel. The FPR is 0 with RBPK for all the four methods. The numbers of SVs are least compared to all the other kernel function, which result into less testing time compared to polynomial and RBF kernel. Figure 10 shows comparison of accuracy and no. of SVs for Letter dataset.

Table 8: Experimental results for Letter Dataset

Methods	Kernel Function	Parameters	Accuracy	Training Time (Sec)	Testing Time (Sec)	Correctly Classified Instances (CCI)	Number of SVs	Precision	TPR	FPR
Holdout (15000/ 5000)	Linear		84.3	7.064	5.07	4215	8770	0.872	0.87	0.002
	Polynomial	$d=3$	37.58	38.454	9.918	1879	14462	0.598	0.387	0.023
	RBF	$\gamma=0.1$	84.6	12.391	9.259	4230	10882	0.882	0.872	0.002
	RBPK	$d=5, \gamma=0.01$	95.88	8.03	5.696	4794	5382	0.988	0.987	0
Training vs. Training (20000/ 20000)	Linear		85.21	11.611	26.058	17042	11165	0.892	0.886	0.002
	Polynomial	$d=3$	43.09	66.108	49.468	8618	19056	0.739	0.444	0.022
	RBF	$\gamma=0.1$	86.5	24.399	52.307	17300	13857	0.905	0.898	0.002
	RBPK	$d=5, \gamma=0.1$	98.655	16.422	33.369	19731	6454	1.028	1.025	0
Cross Validation (k=5)	Linear		83.39	29.381		8339	5183	0.842	0.84	0.001
	Polynomial	$d=3$	26.55	98.552		2655	7831	0.823	0.243	0.005
	RBF	$\gamma=0.1$	81.53	51.884		8153	6444	0.82	0.814	0.001

(10000)	RBPK	d=5, $\gamma=0.1$	93.74	30.347		9374	3515	0.956	0.957	0
Random Sampling (16000/4000)	Linear		83.975	11.3	6.057	3359	9248	0.88	0.872	0.003
	Polynomial	d=3	39.525	55.414	10.049	1581	15378	0.658	0.408	0.023
	RBF	$\gamma=0.1$	84.75	16.513	9.192	3390	11494	0.887	0.88	0.003
	RBPK	d=5, $\gamma=0.08$	95.175	10.49	6.099	3807	5788	0.989	0.989	0

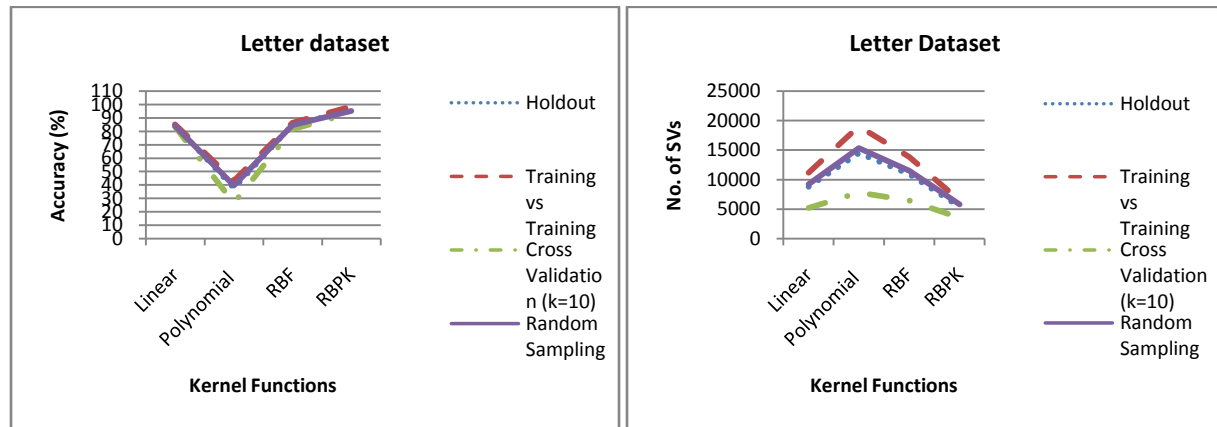


Figure 10: Accuracy and No. of SVs comparison for Letter Dataset

5.7 Result for USPS Dataset

From Table 9, For USPS dataset, the accuracy of RBPK for all four methods is increased about $\approx 0.03\%$ to $\approx 0.5\%$ from the highest accuracy of RBF kernel. The result from the Table 9 shows that the performance of the polynomial kernel is also good for USPS dataset. The precision and TPR is highest for RBPK function compared to other kernel function in all four methods. Figure 11 shows comparison of accuracy and no. of SVs for USPS dataset.

Table 9: Experimental results on USPS Dataset

Methods	Kernel Function	Parameters	Accuracy	Training Time (Sec)	Testing Time (Sec)	Correctly Classified Instances (CCI)	Number of SVs	Precision	TPR	FPR
Holdout (7291/2007)	Linear		93.02	5.906	2.847	1867	992	0.923	0.92	0.008
	Polynomial	d=3	93.77	17.29	6.481	1882	2692	0.926	0.927	0.006
	RBF	$\gamma=0.01$	94.97	11.32	5.02	1833	1833	0.942	0.941	0.005
	RBPK	d=1, $\gamma=0.05$	95.62	15.569	5.42	1919	2029	0.948	0.948	0.003
Training vs. Training (9298/9298)	Linear		99.25	9.645	14.977	9228	1299	1.013	1.012	0
	Polynomial	d=3	97	25.01	35.639	9019	3243	0.99	0.988	0.001
	RBF	$\gamma=0.1$	99.97	141.86	67.106	9295	5820	1.02	1.02	0
	RBPK	d=3, $\gamma=0.01$	100	70.794	48.47	9298	4281	1.02	1.02	0
Cross Validation (k=5) (9298)	Linear		95.43	47.362		8873	1105	0.965	0.963	0.003
	Polynomial	d=3	96.22	126.03		8947	2779	0.972	0.973	0.002
	RBF	$\gamma=0.01$	97.5	82.614		9066	1959	0.983	0.986	0
	RBPK	d=1, $\gamma=0.01$	97.73	56.261		9087	1432	0.996	0.997	0
Random Sampling (6500/2798)	Linear		95	5.295	3.764	2658	1024	0.969	0.969	0.005
	Polynomial	d=3	95.89	14.345	8.438	2683	2542	0.977	0.977	0.004
	RBF	$\gamma=0.05$	97.14	36.32	10.847	2718	3022	0.991	0.99	0.001
	RBPK	d=1, $\gamma=0.05$	97.46	13.294	7.132	2727	2001	0.993	0.994	0.002

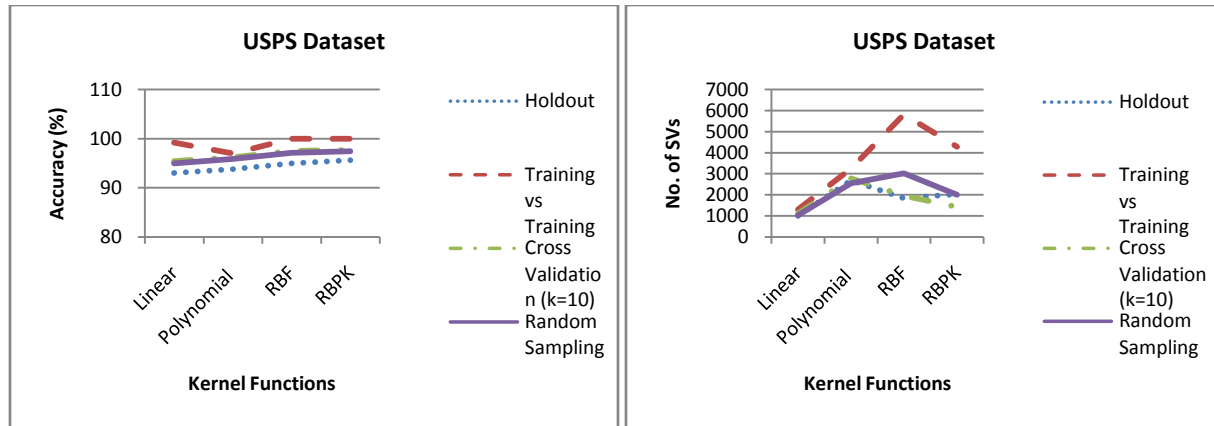


Figure 11: Accuracy and No. of SVs comparison for USPS Dataset

5.8 Result for Web8 Dataset

The promising results are visible in Table 10 for Web8 dataset with the highest accuracy of RBPK. Though the increase of accuracy for different methods is from 0.04% to 0.3% from the highest accuracy of RBF kernel, the number of correctly classified instances is increase up to 40. The web8 dataset is imbalanced dataset, as it contains 97% instances of one class and 3% instances of other class. The performance of the polynomial kernel is around 97%, but it is classifying only one class data correctly, which can be seen from the FPR of it from the Table 10. The numbers of SVs for RBPK is more than linear kernel, but lesser than polynomial and RBF kernel. For web8 dataset the testing time is proportional to number of SVs. Figure 12 shows comparison of accuracy and no. of SVs for Web8 dataset.

Table 10: Experimental results on Web8 Dataset

Methods	Kernel Function	Parameters	Accuracy	Training Time (Sec)	Testing Time (Sec)	Correctly Classified Instances (CCI)	Number of SVs	Precision	TPR	FPR
Holdout (45546/13699)	Linear		98.8	27.704	2.2	13547	1356	0.989	0.989	0.326
	Polynomial	d=3	96.99	18.505	4.82	13288	2678	0.941	0.97	0.97
	RBF	$\gamma=0.08$	99.21	108.159	11.45	13592	3515	0.992	0.992	0.234
	RBPK	d=3, $\gamma=0.05$	99.51	5306.575	5.79	13632	2124	0.995	0.995	0.134
Training vs. Training (45546/45546)	Linear		98.99	27.704	7.182	45086	1356	0.989	0.99	0.291
	Polynomial	d=3	97.1	18.505	16.07	44226	2678	0.942	0.97	0.97
	RBF	$\gamma=0.1$	99.35	149.477	50.438	45249	4279	0.993	0.993	0.195
	RBPK	d=1, $\gamma=0.1$	99.46	67.613	16.033	45298	1956	0.994	0.994	0.155
Cross Validation (k=5) (30000)	Linear		98.9	50.81		29669	779	0.992	0.988	0.077
	Polynomial	d=3	97.1	58.67		29123	1442	1	0.971	0
	RBF	$\gamma=0.08$	98.9	212.76		29672	2239	0.992	0.989	0.038
	RBPK	d=1, $\gamma=0.1$	99.07	67.411		29723	1260	0.993	0.991	0.086
Random Sampling (41470/17775)	Linear		98.9	25.5	3.549	17579	1291	0.988	0.988	0.314
	Polynomial	d=3	97.1	18.96	7.315	17256	2472	0.942	0.97	0.97
	RBF	$\gamma=0.1$	99.13	151.05	22.51	17621	4016	0.991	0.991	0.266
	RBPK	d=1, $\gamma=0.1$	99.17	54.327	6.073	17628	1901	0.991	0.991	0.224

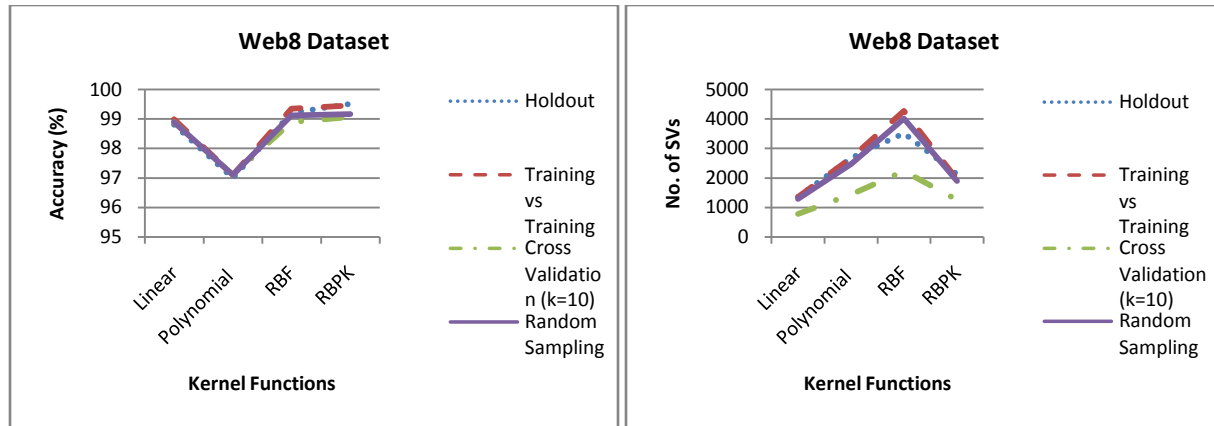


Figure 12: Accuracy and No. of SVs comparison for Web8 Dataset

5.9 Result for Forest Cover Dataset

The Forest Cover Type dataset contains forest cover type data from the US forest service. It is composed of 5,81,102 tuples associated with 30 x 30 meter cells. Each tuples has 54 attributes, of which 10 are quantitative, 4 are binary wilderness areas, and 40 are binary soil type variables. The data are partitioned into seven classes.

Since, the dataset is very large, random sampling is used to select 10000 samples for training and 5000 samples for testing. Since the same training and testing samples are used for holdout method, separate performance is not shown. As seen from Table 11, For Forest Cover type dataset, the accuracy of RBPK for all the three methods is highest. The accuracy for different methods is increased from $\approx 5\%$ to 9% from the highest accuracy of linear kernel. The performance of polynomial kernel is very poor for forest cover type dataset. The number of SVs is least for RBPK for all methods. Figure 13 shows comparison of accuracy and no. of SVs for Web8 dataset.

Table 11: Experimental results for Forest cover Dataset

Methods	Kernel Function	Parameters	Accuracy	Training Time (Sec)	Testing Time (Sec)	Correctly Classified Instances (CCI)	Number of SVs	Precision	TPR	FPR
Training vs. Training (10000/10000)	Linear		77.35	5.609	2.954	7735	5308	0.771	0.773	0.36
	Polynomial	d=3	65.99	8.432	5.249	6599	6802	0.436	0.66	0.66
	RBF	$\gamma=0.1$	74.64	7.768	5.755	7464	5524	0.75	0.746	0.437
	RBPK	d=5, $\gamma=0.08$	86.71	27.433	6.125	8671	4070	0.866	0.867	0.195
Cross Validation (k=5) (10000)	Linear		77.23	34.881		7723	4284	0.824	0.772	0.237
	Polynomial	d=3	65.99	51.644		6599	5440	1	0.66	0
	RBF	$\gamma=0.1$	74.07	44.19		7407	4467	0.855	0.74	0.24
	RBPK	d=5, $\gamma=0.08$	82.58	102.202		8258	3309	0.835	0.825	0.201
Random Sampling (10000/4000)	Linear		78.03	5.609	1.323	3121	5308	0.777	0.78	0.347
	Polynomial	d=3	66	8.432	2.098	2640	6802	0.436	0.66	0.66
	RBF	$\gamma=0.1$	74.28	7.768	2.765	2971	5524	0.744	0.743	0.441
	RBPK	d=5, $\gamma=0.08$	83.55	27.433	2.469	3342	4070	0.833	0.836	0.237

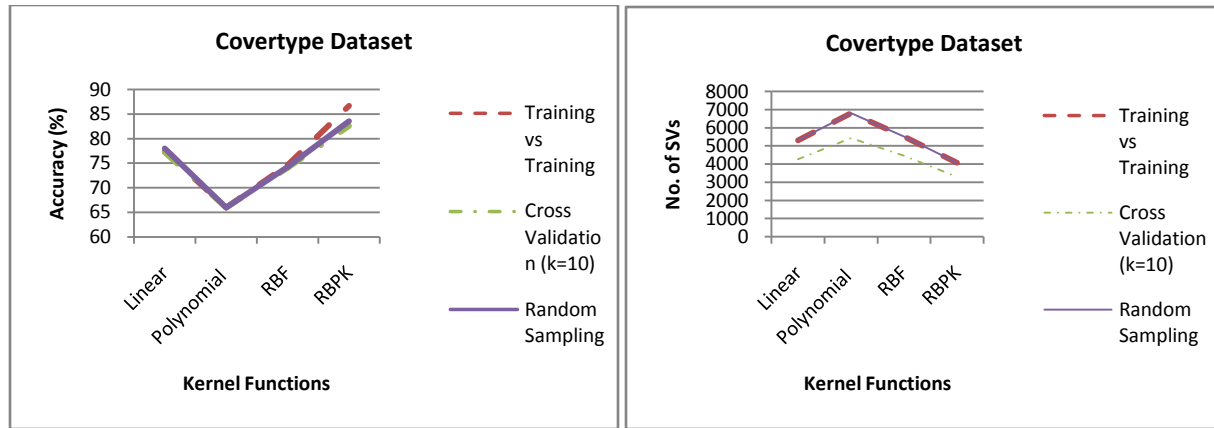


Figure 13: Accuracy and No. of SVs comparison for Coverttype Dataset

6. Conclusion

The proposed kernel function, RBPK, combines the advantages of RBF and Polynomial kernel. Choosing appropriate kernel parameters results into better generalization, learning and predicting capability. Better predicting results can be obtained irrespective of binary or multiclass datasets. The performance of the proposed kernel function, RBPK, is compared with the existing kernel functions named Linear, Polynomial and RBF using mentioned datasets. It clearly shows, from the experiments, that the RBPK has much better accuracy in correctly classifying the instances. For Holdout, Cross-validation and Random sampling method, the accuracy of classifier with RBPK, in correctly classifying instances, increases from 0.03 % to 12 %. For training vs. training method, the accuracy of classifier with RBPK in comparison with other existing kernel function is increased around 0.04% to 25%. The RBPK gives better feature space representation for multiclass classification which ultimately increases the classification accuracy of a classifier. Classification time and Computational complexity for the binary as well as multiclass SVM classifier depend on the number of support vectors required. By mapping the data into new feature space, using RBPK, reduces the number of support vectors compared to the other kernel functions resulting in reduction of the classification time. For the SVM classification, the memory required is directly proportional to the number of support vectors. Hence, support vectors must be reduced to speed up the classification process and to minimize the computational and hardware resources. The number of support vectors is reduced with RBPK, and hence, it requires less memory as compared to the other kernel functions. The promising results on different datasets show that RBPK can be applied to different data domains.

References

- [1] A. Asuncion and D. Newman. UCI machine learning repository, 2007.
- [2] A. Daemen and B. De Moor. "Development of a kernel function for clinical data." In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pp. 5913-5917. IEEE, 2009.
- [3] A. T. Quang, Q. L. Zhang, and X. Li. "Evolving support vector machine parameters." *Machine Learning and Cybernetics, 2002. Proceedings. 2002 International Conference on*. Vol. 1. IEEE, 2002.
- [4] B. Schölkopf, and A. J. Smola, "*Learning with kernels: support vector machines, regularization, optimization, and beyond*". MIT press, 2002.
- [5] C. C. Chang and C. J. Lin. "LIBSVM: a library for support vector machines." *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3): 27, 2011.
- [6] C. Campbell and Y. Ying, "Learning with support vector machines." *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 5(1): 1-95, 2011.
- [7] C. Cortes and V. Vapnik. "Support-vector networks." *Machine learning*, 20(3): 273-297, 1995.
- [8] C. J. Burges, "A tutorial on support vector machines for pattern recognition." *Data mining and knowledge discovery* 2(2), 121-167, 1998
- [9] C. Staelin. "Parameter selection for support vector machines." *Hewlett-Packard Company, Tech. Rep. HPL-2002-354R1* (2003).
- [10] G. F. Smits and E. M. Jordaen. "Improved SVM regression using mixtures of kernels." In *Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on*, vol. 3, pp. 2785-2790. IEEE, 2002.
- [11] G. Mak, "The implementation of support vector machines using the sequential minimal optimization algorithm." PhD diss., McGill University, 2000.
- [12] H. Song, Z. Ding, C. Guo, Z. Li, and H. Xia. "Research on combination kernel function of support vector machine." In *International Conference on Computer Science and Software Engineering, 2008*, vol. 1, pp. 838-841. IEEE, 2008.
- [13] H.Yu and S. Kim. "SVM Tutorial—Classification, Regression and Ranking." In *Handbook of Natural Computing*, pp. 479-506. Springer Berlin Heidelberg, 2012.
- [14] J. C. Platt, "Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines", Microsoft Research 1998
- [15] J. Han and M. Kamber, "Data Mining, Southeast Asia Edition: Concepts and Techniques". Morgan kaufmann, 2006.
- [16] J. Sun. "Fast tuning of SVM kernel parameter using distance between two classes." *Intelligent System and Knowledge Engineering, 2008. ISKE 2008. 3rd International Conference on*. Vol. 1. IEEE, 2008, pp. 108-113.
- [17] K. P. Wu and S. D. Wang. "Choosing the kernel parameters for support vector machines by the inter-cluster distance in the feature space." *Pattern Recognition* 42(5), (2009): 710-717.
- [18] K. R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf. "An introduction to kernel-based learning algorithms." *IEEE Transactions on Neural Networks*, 12(2): 181-201, 2001.
- [19] M. Lu, C. P. Chen, J. Huo, and X. Wang, "Optimization of combined kernel function for svm based on large margin learning theory." In *IEEE International Conference on Systems, Man and Cybernetics, 2008. SMC 2008.*, pp. 353-358. IEEE, 2008.
- [20] M. Tarhouni, K. Laabidi, S. Zidi and M. Ksouri-Lahmari.. A nonlinear MIMO system identification based on improved Multi-Kernel Least Squares Support Vector Machines (Improved Multi-Kernel LS-SVM). In *Systems, Signals and Devices (SSD), 2011 8th International Multi-Conference on* (pp. 1-7). IEEE, 2011.
- [21] R. Burbidge and B. Buxton, "An introduction to support vector machines for data mining." *Keynote papers, young OR12*: 3-15, 2001.
- [22] R. Herbrich. "Learning Kernel classifiers: theory and algorithms (adaptive computation and machine learning)." MIT press, 2001.
- [23] R. Zhang and X. Duan. "A new compositional kernel method for multiple kernels." In *Computer Design and Applications (ICCD), 2010 International Conference on*, vol. 1, pp. V1-27. IEEE, 2010.
- [24] S. Abe. "Training of support vector machines with Mahalanobis kernels." In *Artificial Neural Networks: Formal Models and Their Applications-ICANN 2005*, pp. 571-576. Springer Berlin Heidelberg, 2005.
- [25] S. Guangzhi, D. Lianglong, H. Junchuan, and Z. Yanxia. "Dynamic support vector machine by distributing kernel function."

- In *Advanced Computer Control (ICACC), 2010 2nd International Conference on*, vol. 2, pp. 362-365. IEEE, 2010.
- [26] S. I. Amari and S. Wu. "Improving support vector machine classifiers by modifying kernel functions." *Neural Networks* 12(6): 783-789, 1999.
 - [27] T. Wang. "Improving SVM classification by feature weight learning." *Intelligent Computation Technology and Automation (ICICTA), 2010 International Conference on*. Vol. 2. IEEE, 2010, pp. 518-521.
 - [28] U. Van Luxburg and B. Schölkopf. "Statistical learning theory: Models, concepts, and results." *arXiv preprint arXiv:0810.4752*, 2008.
 - [29] V. N. Vapnik "An overview of statistical learning theory." *Neural Networks, IEEE Transactions on*, 10(5): 988-999, 1999.
 - [30] W. An-na, Z. Yue, H. Yun-tao, and L. Yun-lu, "A novel construction of SVM compound kernel function." *International Conference on Logistics Systems and Intelligent Management*, vol. 3, pp. 1462-1465. IEEE, 2010.
 - [31] X. Mu and Y. Zhou. "A Novel Gaussian Kernel Function for Minimax Probability Machine." In *Intelligent Systems, 2009. GCIS'09. WRI Global Congress on*, vol. 3, pp. 491-494. IEEE, 2009.
 - [32] Y. Jin, J. Huang, and J. Zhang, "Study on influences of model parameters on the performance of SVM." In *International Conference on Electrical and Control Engineering (ICECE)*, pp. 3667-3670. IEEE, 2011.

MOiD (Multiple Objects incremental DBSCAN) – A paradigm shift in incremental DBSCAN

Neha Soni¹, Dr. Amit Ganatra²

¹ Computer Engineering Dept.
SVIT, Vasad, Gujarat, India

² Faculty of Tech. & Engg.,
CHARUSAT, Changa, Gujarat, India

Abstract. Mining an unprecedented increasing volume of data is a herculean task. Many mining techniques are available and being proposed every day. Clustering is one of those techniques used to group unlabeled data. Among prevailing proposed methods of clustering, DBSCAN is a density based clustering method widely used for spatial data. The major problems of DBSCAN algorithm are, its time complexity, handling of varied density datasets, parameter settings etc. Incremental version of DBSCAN has also been proposed to work in dynamic environment but the size of increment is restricted to one data object at a time. This paper presents a new flavour of incremental DBSCAN which works for multiple data objects at a time, named MOiD (Multiple Objects incremental DBSCAN). MOiD has been experimented on thirteen publicly available two dimensional and multi-dimensional datasets. The results show that MOiD performs significantly well in terms of clustering speed with a minor variation in accuracy.

Keywords: Incremental Clustering, DBSCAN, Density based clustering, region query, clustering

1. Introduction

Data Mining is a process of analyzing data from a different perspective and summarizing it into useful information. Today, when the whole world is worried about managing of Big Data, mining of such large amount of data is like finding a needle from

haystack. There are many data mining techniques available like classification, clustering, pattern recognition etc. Each technique has its own merits and demerits with its applicability in certain domain. Based on training of datasets, if it is trained or untrained, these techniques are classified as either supervised or unsupervised mining. Clustering is an unsupervised learning methodology, used for exploration of relationships among the data objects and group them based on its characteristics. Clustering has wide spectrum of applications like, to gain insight to the data distribution, to generate hypotheses, to observe the characteristic and find anomalies, to form the natural classification and even to summarize data. Numerous algorithms have been developed for clustering. Based on working principle, they are broadly categorised as partitioning algorithms, hierarchical, model based and density based [1][2][3][4][5][6].

Partitioning methods, as the name conveys, in general, creates k partitions of the datasets with n objects, each partition represent a cluster, where $k \leq n$. It is assumed that each cluster has at least one object and each object belongs to only one cluster. Partitioning methods try to divide the data into subsets or partitions based on some evaluation criteria. As checking of all possible partitions are computationally infeasible, certain greedy heuristics are used in the form of iterative optimization. The partitioning algorithm in which each cluster is represented by the gravity of the centre is known as k -means algorithms. Since the invention of k -means, large numbers of variations have been proposed, for instance, ISODATA, Forgy, bisecting k -means, x -means, kernel k -means etc [1]. The other type in which a cluster is represented by one of the objects located near its centre is called as a k -medioids. PAM, CLARA and CLARANS are the three main algorithms proposed under the k -medioid method [2].

In the hierarchical methods, the dataset of n objects is decomposed into a hierarchy of groups. This decomposition can be represented by a tree structure diagram called as a dendrogram; whose root node represents the whole dataset and each leaf node is a single object of the dataset. The clustering results can be obtained by cutting the dendrogram at different level. There are two general approaches for the hierarchical method: agglomerative (bottom-up) and divisive (top down). In agglomerative method, it starts with n leaf nodes (n clusters) and in successive steps, it applies merge operation to reach to root node, which would be a cluster containing all data objects. Divisive method works exactly in the reverse order. The merge operation is based on the distance between two clusters. There are three different notions of distance: single link, average link, complete link [3][5].

Model based methods (aka probabilistic models) are used to find the most likely set of clusters for the given data, by assigning a certain probability of object into each cluster. It assumes that the data comes from a mixture of several populations whose distributions and priors are to be calculated [2]. The mixture is a set of k probability distributions, representing k clusters, that governs the attribute values for members of that cluster. The representative algorithms are EM, SNOB, AUTOCLASS and MCLUST [3].

Density based clustering is based on the notion of density where density is considered as the number of objects in the given region. The general idea is to continue growing the

given region as long as the density in the neighbourhood exceeds some threshold; that is for each data point within a given cluster; the neighbourhood of a given radius has to contain at least a minimum number of points. The density based algorithms are classified based on connectivity of points and density functions. The main representative algorithms in the former are DBSCAN and its extensions, OPTICS, DBCLASD, whereas, DENCLUE and SNN [2][3] are for the later.

Recent advances in storage, network and computer technology and massive use of smart devices, result in producing huge data volumes. For instance, Facebook collects 500TB data per day. It is computationally expensive to analyze clustering structures in such a massive data where the data is dynamic. Static clustering is not cost effective in such environment. This gave birth to the idea of incremental clustering algorithms. In incremental clustering, the new data points are directly added to the existing clusters by applying incremental clustering algorithm. This makes it cost effective in terms of time and space as well. The overall idea of incremental clustering can be visualized as given in figure 1.

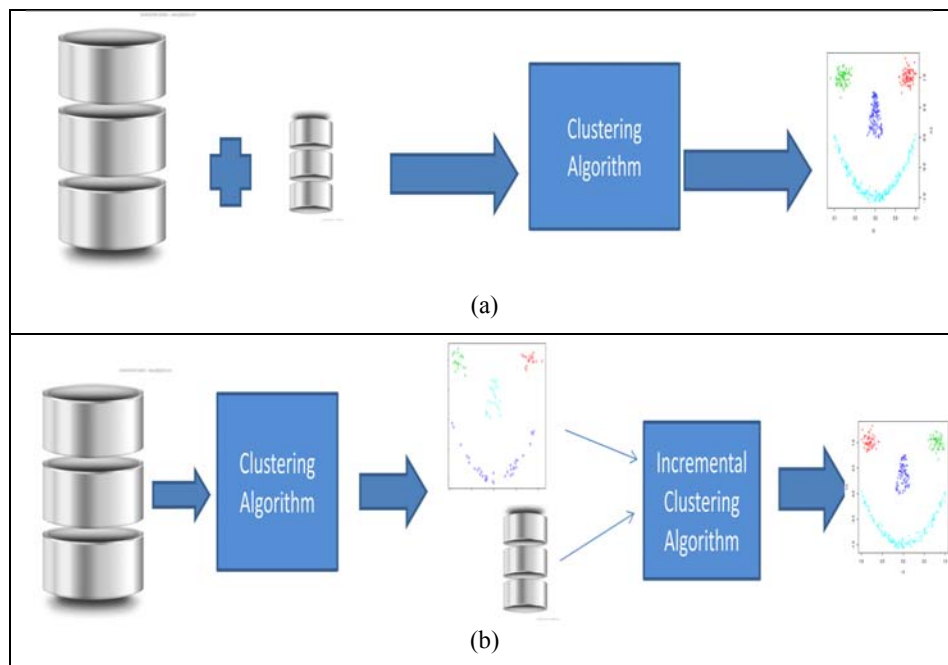


Fig. 1. (a) Static Clustering approach (b) Incremental Clustering approach

Many incremental clustering algorithms have been proposed to handle large datasets dynamically. The major advantage of the incremental approach is their limited space requirement since the entire dataset is not necessary to be stored in the memory. Therefore, these algorithms are well suited for a dynamic environment and for very large datasets [20].

The next section of the paper discusses about the related work in the area of clustering, mainly on DBSCAN. It also throws light on work done on other incremental approaches. Section 3 explains the proposed approach and section 4 contains a detailed discussion on the experimental results. Finally, the paper ends with conclusion and references about the invaluable contribution of the authors.

2. Related Work

Many flavours of DBSCAN have been proposed by the researchers. Incremental approach has also been proposed for various other data mining techniques. This section goes through the work of various authors on DBSCAN and incremental approaches.

FDBSCAN, in [8], tries to reduce the time complexity by reducing the number of region query. DBSCAN carries out region query operation for every object contained in the core object's neighbourhood. Here, in FDBSCAN, instead of selecting all the objects in core point p 's neighborhood, only some representative objects are considered to perform region query. Two different algorithms have been also proposed for the selection of representative objects. The l-DBSCAN [9] proposes that two types of prototype can only be used to execute region query, instead of all points in the core point's neighbourhood. The two types of prototypes are: at a coarser level to reduce the time requirement and at finer level to improve the deviated result. Prototypes are derived using leaders clustering method. It is suitable for large data sets. The time complexity of l-DBSCAN is $O(n+k^2)$ where k is the number of leaders. In MEDBSCAN[10], a new parameter MDV(Max. Distance Value) is added and the region query is performed for only those objects which fall between the range of ϵ -neighbourhood and MDV, thereby reducing the number of region queries which results into the reduction in time. The ODBSCAN, proposed by J. H. Peter et. al. [11] is based on the combined approach of FDBSCAN and MEDBSCAN. ODBSCAN tries to remove the problem of MEDBSCAN by not selecting only 4 points as representative points for the cluster expansion as in MEDBSCAN. Instead, all border objects are considered in cluster expansion. Another modification proposed by the authors is to maintain two separate queues named InnerRegionObjects queue for holding the points lying between ϵ -radius and OuterRegionObjects queue to hold the points falling between $2*\epsilon$ and ϵ radius. These two queues can all be separated into four different queues to minimize the unwanted distance computation while processing the border objects.

The authors of [12] give a new density function which is based on the kNN-stratification and Influence function. The input dataset is stratified based on the density function and partitioned into layers of decreasing density based on the average k -adjusted influence function and lastly the outliers are heuristically detected. The knowledge inferred during the outlier detection phase is embedded into a new residual space by adding the new dimension. Each value in this dimension represents the sum of the distances of the influence space. It improves the separation of clusters with different densities. Then a modified algorithm is applied to the new residual dataset. The work in [13] is based on the grid partition technique and multi-density based clustering. The authors have proposed the technique for automatic generation of Eps and Minpts parameters of the DBSCAN algorithm. It builds a unified grid size to divide the data space and then stores its internal data statistics in each grid. All the clustering operations

are targeted to the grid cell in order to form the cluster in the integral structure of grid. In [14], the enhanced DBSCAN algorithm has been proposed to handle the datasets of varying densities. The idea is to use different values of Eps according to the local density of the starting point in each cluster. The clustering process starts from the highest to lowest local density point. For each value of Eps, DBSCAN is adopted to make sure that all density reachable points with respect to current Eps are clustered. As a next step, the clustered points are ignored, to avoid merging among the dense clusters with the sparse ones. In [15], the authors propose the algorithm which merges between hierarchical, partition and density based methods. It provides outlier detection and data clustering simultaneously. The algorithm consists of two phase: in the first phase, it removes the outliers from the input dataset, and later, in the second phase, it performs the clustering process. It requires only one parameter as a threshold for the outlier detection. It builds k nearest neighbours graph, estimates the density for each point from this graph and removes the low density points according to the input parameter. After outlier removal, the algorithm starts clustering from the densest point, applying the idea of single link algorithm with some modification to overcome the computational complexity. To construct the k nearest neighbour graph, the idea of canopy is used.

DDSC [16] is an extension of the DBSCAN algorithm to detect the clusters with differing densities. The adjacent regions are separated into different clusters if there is a significant change in densities. It starts a clustering with a homogeneous core object and goes on expanding it by including the other directly density-reachable homogeneous core objects until non-homogeneous core objects are detected. In [17], to determine the parameters Eps and MinPts, the behaviour of the distance from a point to its kth nearest neighbour is checked, which is called k-dist. First of all k-dist are computed for each point for some k value. Then the values are sorted. Then the k-dist graph is drawn using these values. The sharp change visible in graph corresponds to a suitable value of Epsi. The STDBSCAN [18] proposes a new concept where each cluster is assigned a density factor, which is the degree of the density of the cluster. If cluster C is a "loose" cluster, density_distance_min would increase and therefore the density_distance would be quite small. This forces the density factor of C to be quite close to 1. The scenario would be reversed if C is a "tight" cluster with density factor value close to 0. Locally Scaled Density Based Clustering [19] introduces a notion of local scaling which determines the density threshold based on the local statistics of the data. The local maxima of density are discovered using a k-nearest-neighbour density estimation and are used as centres of potential clusters. Each cluster is grown until the density falls below a pre-specified ratio of the centre point's density. This makes the clustering more robust and does not require fine tuning of parameters. The algorithm needs two parameters: firstly, k- the order of nearest neighbour to consider and secondly, an α to decide when the drop in the density is necessary for the cluster change. The computational complexity of the proposed algorithm is same as that of the DBSCAN.

In clustering community, apart from the various flavours of DBSCAN, the work has been carried out on incremental clustering approach too over the decades. The study of incremental clustering dates back to late 70's in the last century. In [20], a dynamic and incremental clustering algorithm is proposed, based on density-reachable criteria, to improve the efficiency of data resource utilization as well as processes clusters of arbitrary shapes with noise or outliers. The situation, where the constraints are

incremental, is exploited in [21]. Incremental Model-Based Clustering for Large Datasets with Small Clusters is proposed in [22]. The method starts by drawing a random sample of the data, selecting and fitting a clustering model to the sample, and extending the model to the full dataset by additional Expectation Maximization iterations. New clusters are then added incrementally. Incremental Clustering for Dynamic Information Processing proposed in [23], called C2ICM, creates an effective and efficient retrieval environment, and has been proven cost effective with respect to re-clustering. In [24], by C. C. Hsu et. al., a MART algorithm that can handle mixed dataset directly is presented. It introduces the distance hierarchy tree structure to overcome the expression for similar degree. This distance hierarchy tree algorithm combines the adaptive resonance theory network algorithm and it can be effective with mixed data in clustering. In [25], the authors have used a compact representation of a mobile trajectory and defined a new similarity measure between the trajectories. An incremental clustering algorithm for finding the evolving groups of similar mobile objects in spatio-temporal data is also proposed.

[26] proposes a distance based incremental clustering method which finds the clusters of arbitrary shapes and sizes in fast changing databases. It is based on scalable distance based clustering al-SL, which finds arbitrary shaped clusters in metric databases. The search space is restricted for potential changes of clustering membership of patterns after inclusion of new points in the dataset using leaders clustering and metric space properties. The proposed incremental clustering method IncrementalSL is based on distance based al-SL method.

The [27] proposes a model-called incremental clustering which is based on a careful analysis of the requirements of the information retrieval application, and useful in other applications. The goal is to efficiently maintain clusters of small diameter as new points are inserted. The model enforces the requirement that all times an incremental algorithm should maintain a HAC for the points presented up to that time. The algorithm is free to use any rule for choosing the two clusters to merge at each step. This model preserves all the desirable properties of HAC while providing a clean extension to the dynamic case. In addition, it has been observed that such incremental algorithms exhibit good paging performance when the clusters themselves are stored in secondary storage, while cluster representatives are preserved in main memory.

The GRIN algorithm [28], an incremental hierarchical clustering algorithm for the numerical datasets based on gravity theory in physics. The GRIN algorithm delivers favorite clustering quality and generally features $O(n)$ time complexity, for, the optimal parameters settings in the GRIN algorithm are not sensitive to the distribution of the data set. The GRIN algorithm operates in two phases. In both phases, it invokes the gravity-based agglomerative hierarchical clustering algorithm to construct clustering dendrograms. In the first phase of the algorithm, a number of samples are taken from the incoming data pool and the GRACE algorithm is invoked to build a clustering dendrogram for these samples. The clustering quality of the GRIN algorithm is immune from how incoming data instances are ordered. However, the order of incoming data instances may impact the execution time of the second phase of the GRIN algorithm.

The thorough analysis infers that there is no algorithm available based on density based clustering with incremental approach working on multiple data objects simultaneously. This motivated the incubation of MOiD.

3. MOiD – Multiple Objects incremental DBSCAN

The proposed work is based on the concepts of incremental DBSCAN. As stated earlier, in contrast to incremental DBSCAN [30] which works for a single object at a time, this approach works and can handle multiple objects. Hence it is named as Multiple Objects incremental DBSCAN. Before peeping into the proposed work, since MOiD is based on DBSCAN and incremental DBSCAN approach, the following text summarizes both.

3.1 DBSCAN

The basic idea of density based clustering involves a number of new terminologies as defined below with example shown in figure 2 [2] [29].

1. ϵ -neighbourhood: The neighbourhood within a radius ϵ of a given object is called the ϵ -neighbourhood of the object.
2. core object: If the ϵ -neighbourhood of an object contains at least a minimum number, MinPts, of objects, then the object is called a core object.
3. border point: A border point has fewer than MinPts within radius ϵ , but is in the neighborhood of a core point.
4. directly density-reachable: given a set of objects D, an object p is directly density-reachable from object q if p is within the ϵ -neighbourhood of q, and q is a core object.
5. (indirectly) density-reachable: an object p is density-reachable from object q w.r.t ϵ and MinPts in a set of objects, D, if there is a chain of objects p_1, \dots, p_n , where $p_1 = p$ and $p_n = q$ such that p_{i+1} is directly density-reachable from p_i w.r.t ϵ and MinPts, for $1 \leq i \leq n$.
6. density-connected: an object is density-connected to object q w.r.t ϵ and MinPts in a set of objects, D, if there is an object o in D such that both p and q are density-reachable from o w.r.t ϵ and MinPts.

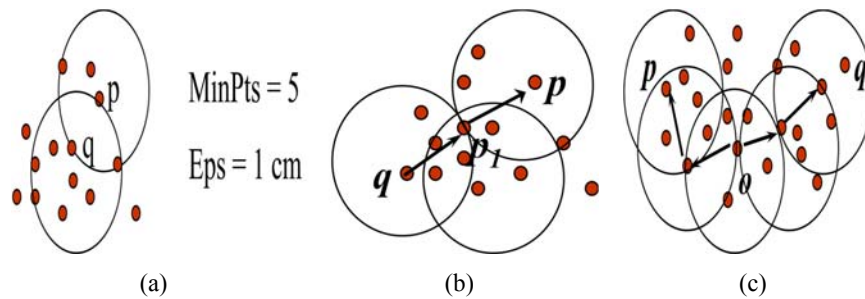


Fig. 2. (a) q core point and p border point (b) Density reachability (c) Density connectivity

The density based algorithms discovers clusters with arbitrary shapes, can handle noise and are more efficient than partitioning and hierarchical methods. However, the first and foremost major problem with the general density based methods is, their time complexity, in general is $O(n^2)$, which is very expensive for the large dataset[2][3]. Second major problem is that they cannot handle varied density datasets and the third, the efficiency of DBSCAN is dependent on the number of different input parameter. If users could not choose good value any of these parameters, this may result in a large number of very small clusters or few very large clusters. To make DBSCAN work in the dynamic environment, an incremental approach has also been proposed. The next subsection explains the approach briefly.

3.2 Incremental DBSCAN

The very first incremental clustering algorithm based on DBSCAN has been proposed for mining in a data warehousing environment by M. Ester et. al. [26]. DBSCAN is based on the density concept and hence it has been proved that the insertion or deletion of single object affects the current cluster only in a small neighbourhood of the object. Based on this argument, the concept of affected objects is given. On an insertion or deletion of an object p , the set of affected objects, i.e. objects which may potentially change the cluster membership after the update, is the set of objects in Eps -neighbourhood of p along with all the density-reachable objects from one of these objects in $D \cup \{p\}$. The cluster membership of all other objects not in the set of affected objects will not change. The definition for the same is as follows:

Let D be a database of objects and p be some object (either in or not in D). The set of objects in D affected by the insertion or deletion of p is

$$\text{Affected}_D(p) = N_{\text{Eps}}(p) \cup \{q \mid \exists o \in N_{\text{Eps}}(p) \wedge q >_{D \cup \{p\}} o\}$$

In incremental DBSCAN, after the insertion or deletion of an object p , the DBSCAN is re-applied on the set of affected objects so as to order to update the clustering.

3.3 MOiD (Multiple Objects incremental DBSCAN)

In order to address the limitations of incremental DBSCAN which only works for single object at a time, this section proposes an incremental approach, MOiD, with the capacity to work on multiple objects simultaneously. It works in two phases. In phase one, it calls DBSCAN to perform the cluster analysis of the incremental (new) dataset. After cluster analysis, MOiD tries to incorporate the clusters of incremental dataset to that of existing dataset. For the understanding purpose, when the new data points are added to the existing dataset, they all together is termed as the New Dataset and the previous data set is referred as the Old Dataset. Based on the proximity of the new data points to be added into the old data set, the working of MOiD is divided into three different scenarios.

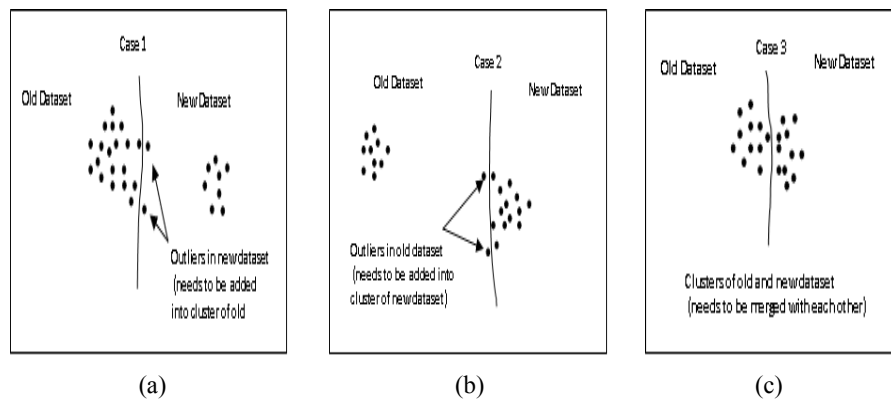


Fig. 3. Representing 3 different scenarios when new dataset clusters are to be incorporated to old dataset clusters.

Case 1: In figure 3(a), when DBSCAN is applied over the New Dataset, few points may be clustered as outliers as they do not fall into any of the cluster. These outliers may be a part of existing cluster of Old Dataset.

Case 2: The second case is opposite to the Case 1. It may happen that, few points which were clustered as the outliers in the Old Dataset, but when New Dataset is added, will become a part of cluster in the New Dataset. This scenario is shown in figure 3(b).

Case 3: The third situation which may arise when the clusters generated in the New Dataset are part of the cluster in the Old Dataset. Figure 3(c) shows the possible scenario.

Taking into the aforementioned cases in consideration, the algorithm has been designed in the following way for each case:

<p>CASE – 1 : points of new dataset classified as a noise/outlier may be part of cluster of old dataset</p> <p>for each cluster co in old dataset Do</p> <p> for each noise point on in new dataset Dn</p> <p> get the eps-neighbourhood points of on in co</p> <p> if number of eps-neighbourhood points > 2 then</p> <p> change cluster label of on to the cluster label of co</p> <p> else</p> <p> mark point on as a noise point.</p>
<p>// CASE – 2 : points of old dataset classified as a noise/outlier may be part of cluster of old dataset</p> <p>for each cluster cn in new dataset Dn</p> <p> for each noise point oo in old dataset Do</p> <p> get the eps-neighbourhood points of oo in cn</p> <p> if number of eps-neighbourhood points > 2 then</p> <p> change cluster label of oo to the cluster label of cn .</p> <p> else</p> <p> mark point oo as a noise point</p>
<p>// CASE – 3: cluster of old dataset and new dataset may merge with each other</p> <p>for each cluster cn in new dataset</p> <p> for each point in pn</p> <p> get the eps-neighbourhood points of pn in co</p> <p> if number of eps-neighbourhood points > 2 then</p> <p> { merge cluster</p> <p> Break</p> <p> }</p>

The time complexity of DBSCAN is $O(N^2)$ without using R^* index [24]. For incremental dataset, if we apply DBSCAN using static approach then the complexity will be $O(N_{o+n}^2)$, where N_{o+n} is the total number of objects in the old dataset plus new dataset. The time complexity of the proposed algorithm is $O(N_n^2) + O(ko * po * kn * pn)$, where N_n is the number of data objects in the new dataset, ko is the number of clusters in the old dataset and kn is number of clusters in the new dataset. po and pn are the current points in the consideration in the old and new datasets respectively. Here, the

value $k_o * p_o * k_n * p_n$ is negligible with respect to N_n and naturally, $N_{n+o} > N_n$. This proves that MOiD performs better than the static DBSCAN.

4 Experiments' Results and Discussions

MOiD performs better than the DBSCAN, as stated in the previous section. To prove this, many data sets have been taken in to the consideration and are tested. Since for the single object incremental DBSCAN, it could not be made possible to test, this section provides an in-depth analysis and discussions on the tests performed on MOiD and static DBSCAN only. In order to prove the robustness of MOiD, two and multidimensional datasets have been used for testing.

4.1 Experimental Setup and Preliminaries

To analyze the performance of the proposed algorithm, six synthetic two dimensional datasets and seven multidimensional data sets are used ranging from 150 to 5000 data points. All the datasets contain the clusters of arbitrary shapes and size. The two parameters required by the DBSCAN, namely Eps and Minpts were tuned for each of the dataset by repeating the cluster analysis using DBSCAN and by comparing the results with the available class labels. The details of all the datasets with the tuned input parameters is shown in table 1.

Table 1: Dataset Details

Dataset Name	Actual Size	No. of Dimensions	Eps	Minpts
Smiley	500	2	0.5	4
R15	600		0.34	4
Aggregation	788		1.5	10
D31	3100		0.65	20
S1	5000		45000	50
S2	5000		35900	50
IRIS	150	4	0.4	4
Wine	178	13	55	5
Heart	270	13	45	4
e-Coli	336	7	0.134	4
AU500	500	125	34000	4
Banknote	1372	4	2	4
EEG	3543	14	42	6

The incremental process for the new data point into the existing dataset is performed in the form of insertion operations, leading to three cases as explained in the previous section. To evaluate the performance of the proposed work for all the three cases, in order to generate the increments, two different methods are used for the selection of the data points.

In the first method, from each dataset, numbers of data points are selected using random sampling method without replacement. The sampled points are then removed from the actual dataset. DBSCAN is applied over it and the clusters are generated. Now, this dataset for which the clusters are generated is designated as Old Dataset. The sampled points are then divided into five different parts and DBSCAN is applied again to generate the clusters. The generated clusters of each different increment are now added gradually to the Old Dataset applying the MOiD.

The second method uses the partitioning concept, in which some fixed numbers of points are removed from the actual datasets in a sequence. DBSCAN is applied over it and the clusters are generated. The Old Dataset is now ready. The removed data points are then divided into five different partitions and DBSCAN is again applied over the partitions to generate the clusters. The generated clusters of each different partition are added to the Old Dataset using MOiD.

The proposed algorithm is implemented in R3.1.2. For DBSCAN implementation, fpc package of R is used [34][35]. All experiments are carried out on HP Pavilion g4 notebook PC, with Intel Core i5 Processor of 2.67 GHz on 02nd November, 2015.

To measure the performance, four different validity indices, execution time and region queries (finding the EPS neighbourhood) are used. These terms are described below.

- The Rand index or Rand measure in data clustering, is a measure of the similarity between two data clustering's[36]. A form of the Rand index may be defined that is adjusted for the chance grouping of elements; this is the adjusted Rand index or corrected Rand Index[37]. From a mathematical standpoint, Rand index is related to the accuracy, but is applicable even when class labels are not used. Given a set of n elements $S = \{o_1, o_2, \dots, o_n\}$ and two partitions of S to compare, $X = \{X_1, X_2, \dots, X_r\}$, a partition of S into r subsets, and $Y = \{Y_1, Y_2, \dots, Y_s\}$, a partition of S into s subsets, define the following:

$$R = \frac{a + b}{a + b + c + d}$$

Where a is the number of pairs of elements in S that are in the same set in X and in the same set in Y ; b is the number of pairs of elements in S that are in the different sets in X and in the different sets in Y ; c is the number of pairs of elements in S that are in the same set in X and in the different sets in Y and d is the number of pairs of elements in S that are in the different set in X and in the same set in Y .

The Rand index has a value between 0 and 1, with 0 indicating that the two data clusters do not agree on any pair of points and 1 indicating that the data clusters are exactly the same.

The adjusted Rand index is the corrected-for-chance version of the Rand index. Though the Rand Index may only yield a value between 0 and +1, the Adjusted Rand Index can yield negative values if the index is less than the expected index.

- Entropy, in [7], is the degree to which each cluster consists of objects of a single class. For each cluster, the class distribution of the data is calculated by $p_{ij} = \frac{m_{ij}}{m_i}$, where m_i is the number of objects in cluster i , m_{ij} is the number of objects of class j in cluster i , and p_{ij} denotes the probability of a member of cluster i belongs the class j . The entropy of each cluster i is calculated using the standard formula:

$$e_i = - \sum_{j=1}^L p_{ij} \log_2 p_{ij}$$

Here L is the number of class. The total entropy for a set of clusters is calculated as:

$$e = - \sum_{j=1}^K \frac{m_i}{m} e_i$$

Where K is the number of clusters and m is the total number of data points.

- Dunn Index [38] is based on the minimum pair wise distance between objects in different clusters as the inter-cluster separation and the maximum diameter among all clusters as the intra-cluster compactness. The larger value of Dunn Index means better cluster configuration.

$$\min_i \left\{ \min_j \left(\frac{\min_{x \in C_i, y \in C_j} d(x, y)}{\max_k \{ \max_{x, y \in C_k} d(x, y) \}} \right) \right\}$$

Where C_i is the i -th cluster, n_i is the number of objects in C_i and $d(x, y)$ is the distance between x and y .

- The silhouette width [39] is the average of each observation's silhouette value. The silhouette value measures the degree of confidence in the clustering assignment of a particular observation, with well-clustered observations having values near 1 and poorly clustered observations having values near -1. For observation i , it is defined as

$$S(i) = \frac{b_i - a_i}{\max(b_i, a_i)}$$

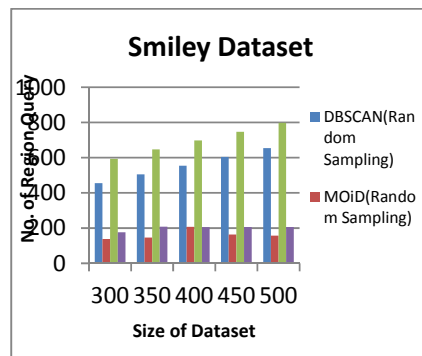
Where a_i is the average distance between i and all other observations in the same cluster, and b_i is the average distance between i and the observations in the nearest neighbouring cluster, i.e.

$$a_i = \frac{1}{n(C(i))} \sum_{j \in C(i)} \text{dist}(i, j), \quad b_i = \min_{C_k \in C \setminus C(i)} \sum_{j \in C_k} \frac{\text{dist}(i, j)}{n(C_k)}$$

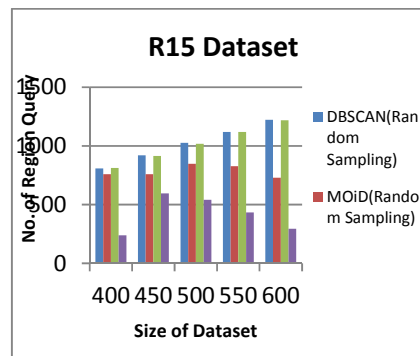
where $C(i)$ is the cluster containing observation i , $\text{dist}(i, j)$ is the distance (e.g. Euclidean, Manhattan) between observations i and j , and $n(C)$ is the cardinality of cluster C . The silhouette width thus lies in the interval $[-1, 1]$, and should be maximized.

4.2 Result Analysis

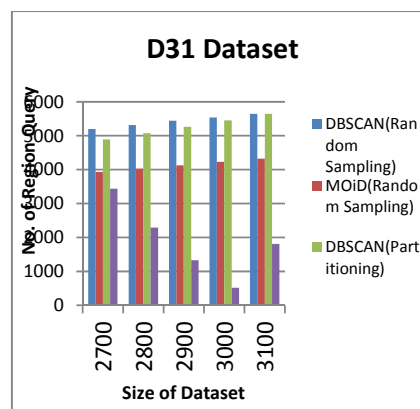
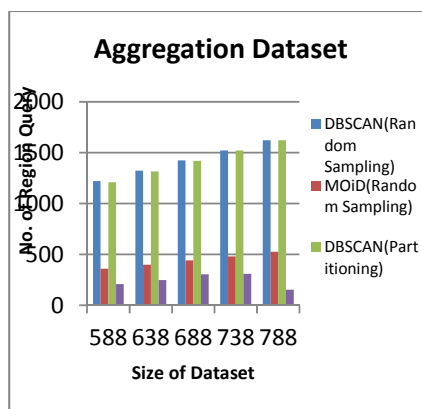
As stated previously that to prove the effectiveness of MOiD, the experiments are performed on the different types of datasets. The names of the two dimensional datasets are Smiley, R15 [31], Aggregation [32], D31, S1 and S2 [33] and the multidimensional datasets are IRIS, Wine, Heart, e-Coli, AU-500, Banknote and EEG [40].



(a)



(b)



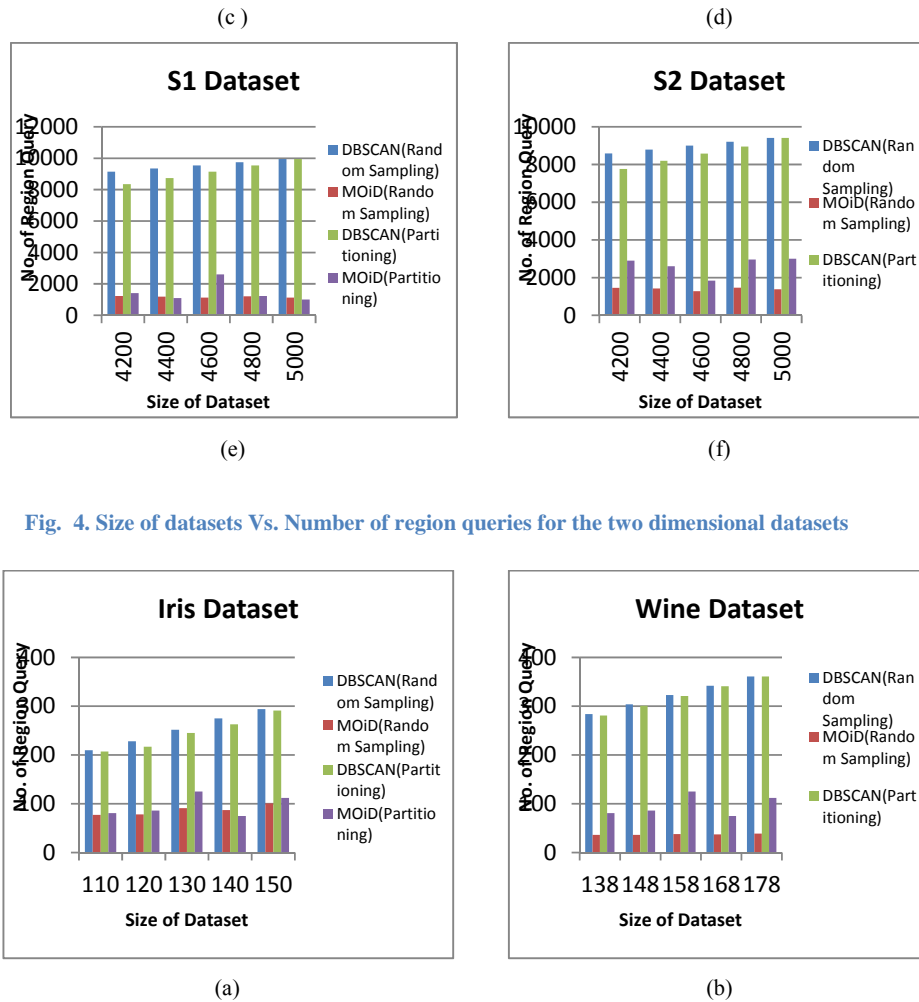
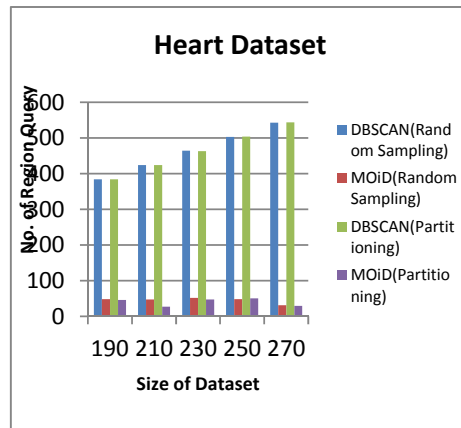
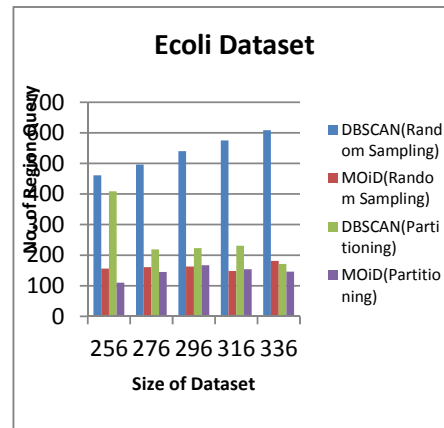


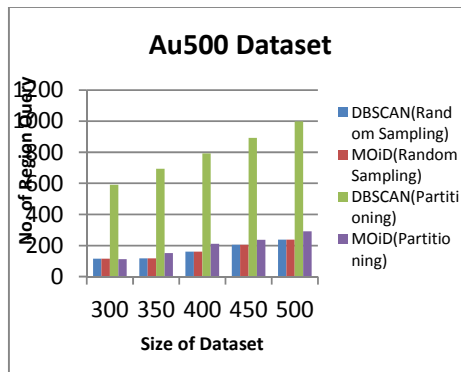
Fig. 4. Size of datasets Vs. Number of region queries for the two dimensional datasets



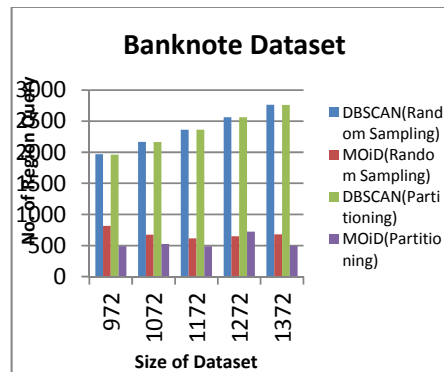
(c)



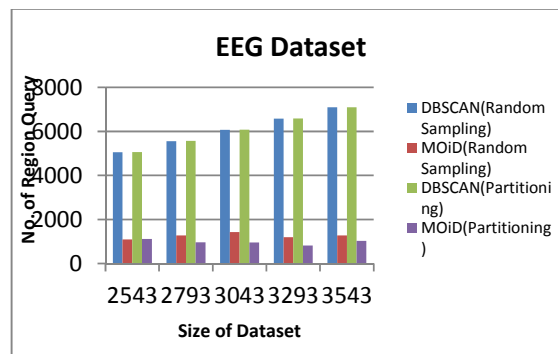
(d)



(e)



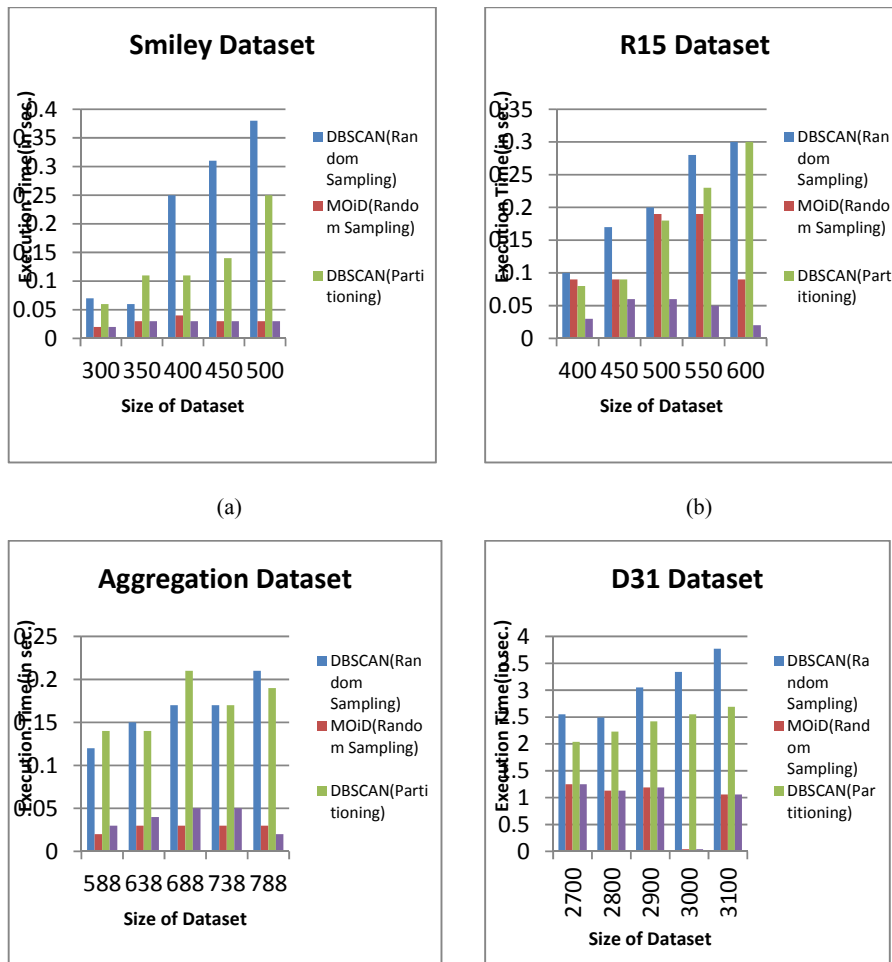
(f)



(g)

Fig. 5. Size of datasets Vs. Number of region queries for the multidimensional datasets

Figure 4 (a) to (f) and figure 5 (a) to (g) shows the number of regions queries against the size of datasets. The size of datasets varies from 500 to 5000 data objects for the two dimensional datasets and the same varies from 150 to 3500 for the multidimensional datasets. For both the kinds of datasets, each size of datasets is sampled based on random and partition sampling. Then the DBSCAN and MOiD are applied on them with different increments. The results prove that MOiD heavily outperforms DBSCAN in both types of sampling for all the datasets. This leads to major reduction in time too.



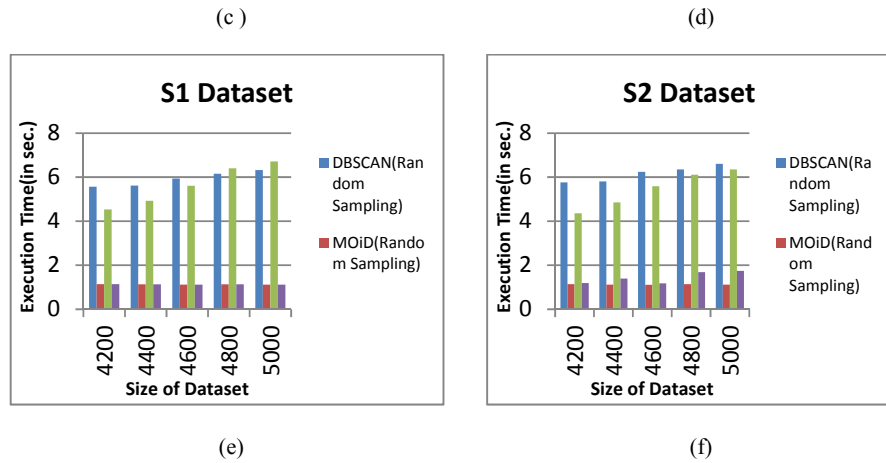
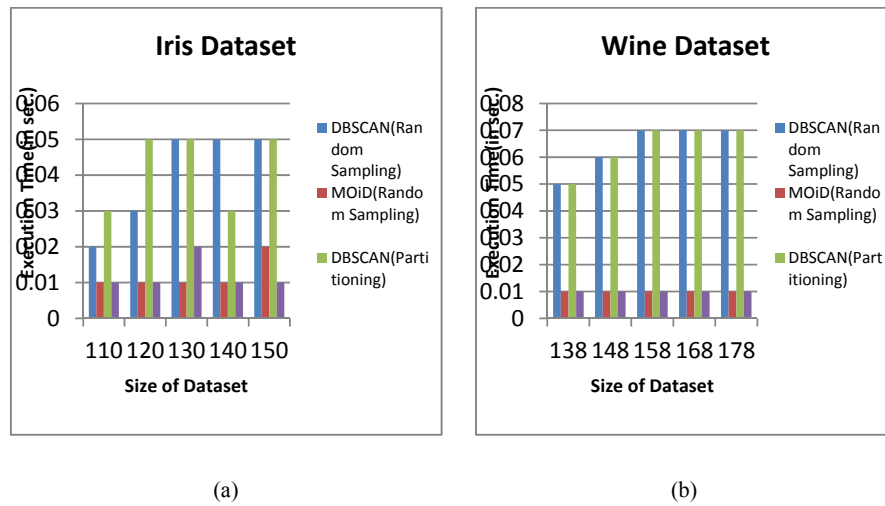
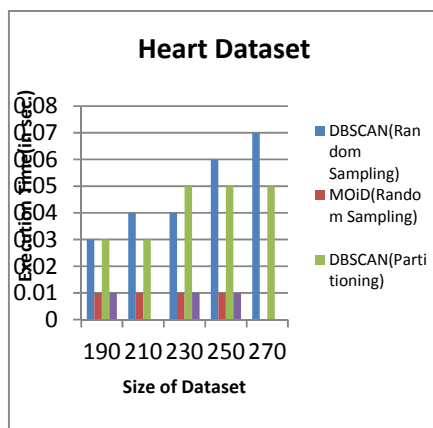
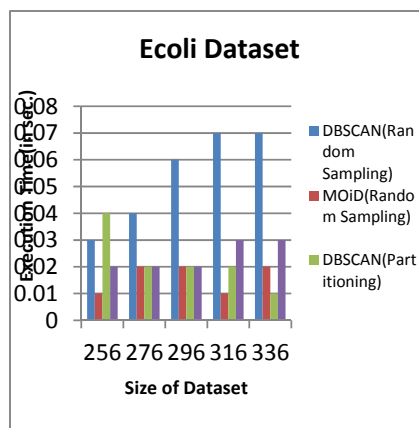


Fig. 6. Size of datasets Vs. Execution time for the two dimensional datasets

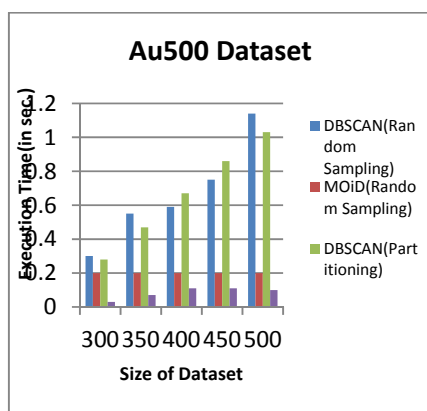




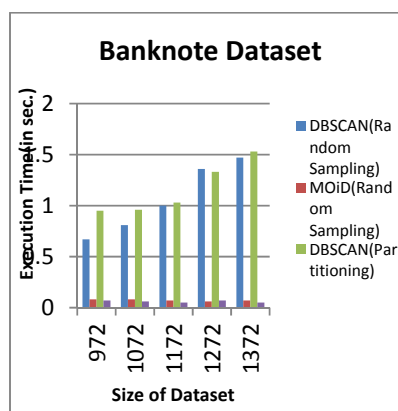
(c)



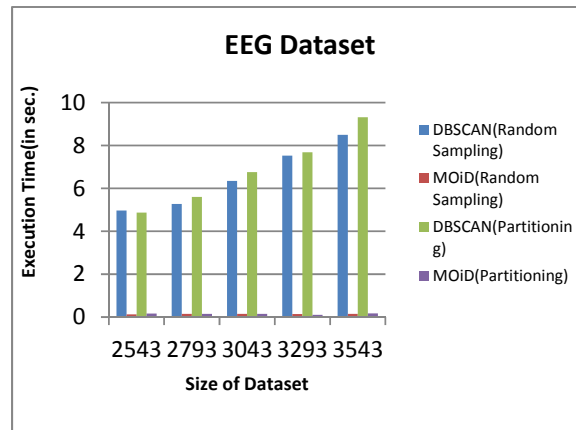
(d)



(e)



(f)



(g)

Fig. 7. Size of datasets Vs. Execution time for the multidimensional datasets

The execution time taken by DBSCAN and MOiD for two and multidimensional datasets is shown in Figure 6 (a) to (f) and figure 7 (a) to (g). The MOiD has the lowest execution time in all the cases. Except e-coli and IRIS, the execution time remains consistent for all increments too. The time taken by MOiD is significantly low with a wide margin, proves that MOiD is highly time efficient as compared to its elder sibling DBSCAN.

The other performance measures derived from the experiments for both DBSCAN and MOiD are shown from figure 8 to 12. In figure 8, the accuracy of both the algorithms is shown. For the MOiD, random and partition sampling are taken into consideration. The results depict overall best performance of MOiD.

The corrected rand index, shown in figure 9, ranges from -1 to 1. The algorithm is said to be good if the values remains closer to 1. In the experiments, except the Heart dataset, for all the datasets, the values are closer to 1 proving the goodness of the algorithm. For Smiley, R15, aggregation, S1 and S2, the rand index is almost 1.

The lower the entropy value, the better the algorithm is. With lower values for all the datasets, except AU500 and EEG, the algorithm shows almost same performance as the DBSCAN. This has been shown in figure 10. The noticeable thing is, for the Heart dataset, MOiD is getting the lowest entropy.

The Dunn index is measured to show the compactness of the clusters. Figure 11 contains values ranging from 0.00 for EEG to 0.5 for Heart dataset. Only in case of Heart, AU500 and Banknote, the DBSCAN performs better over MOiD. For the rest, MOiD performs better.

Silhouette shows how better clustering is performed for the given datasets. Figure 12 shows graph of silhouette values ranging from 0.04 to 0.7. Similar to the Dunn index, heart is having the highest value of 0.7596. For the rest of the datasets, MOiD performed same or better over DBSCAN.

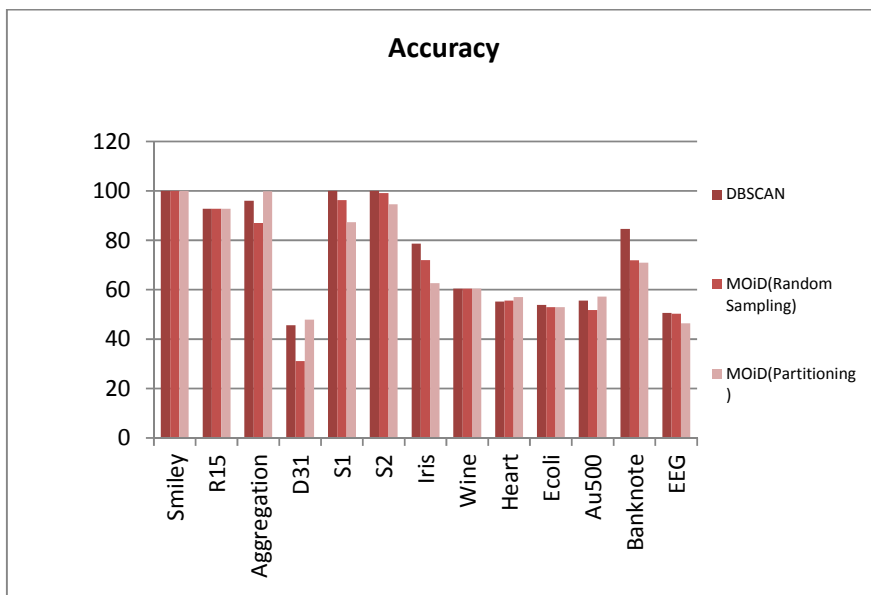


Fig. 8. Datasets Vs. Accuracy of MOiD and DBSCAN

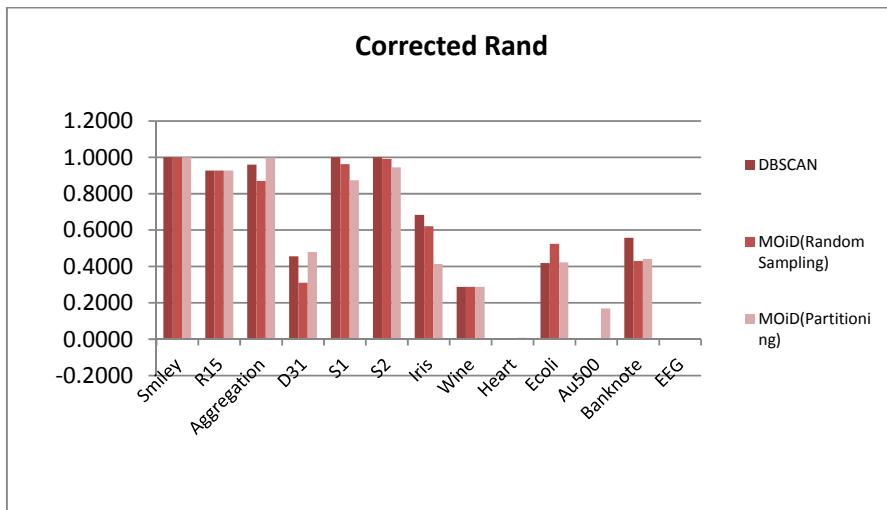


Fig. 9. Datasets Vs. Corrected Rand of MOiD and DBSCAN

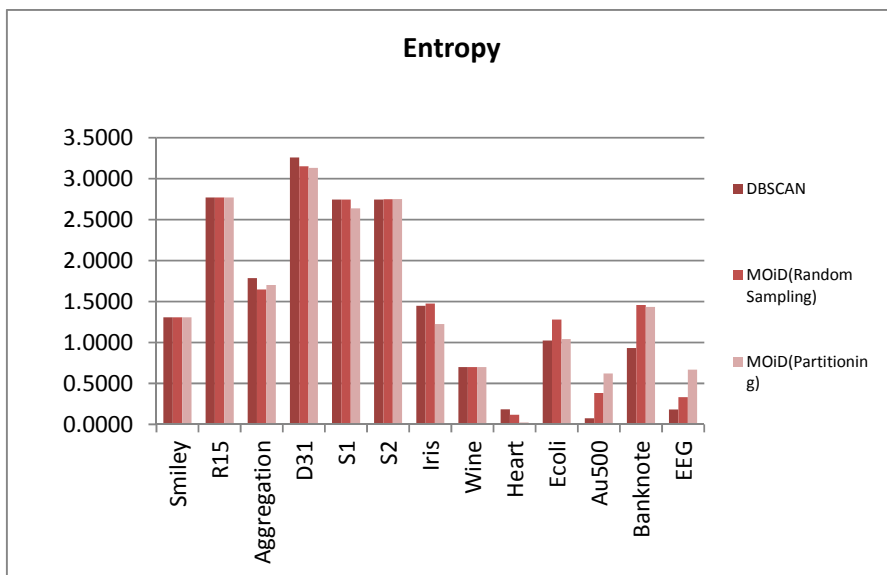


Fig. 10. Datasets Vs. Entropy of MOiD and DBSCAN

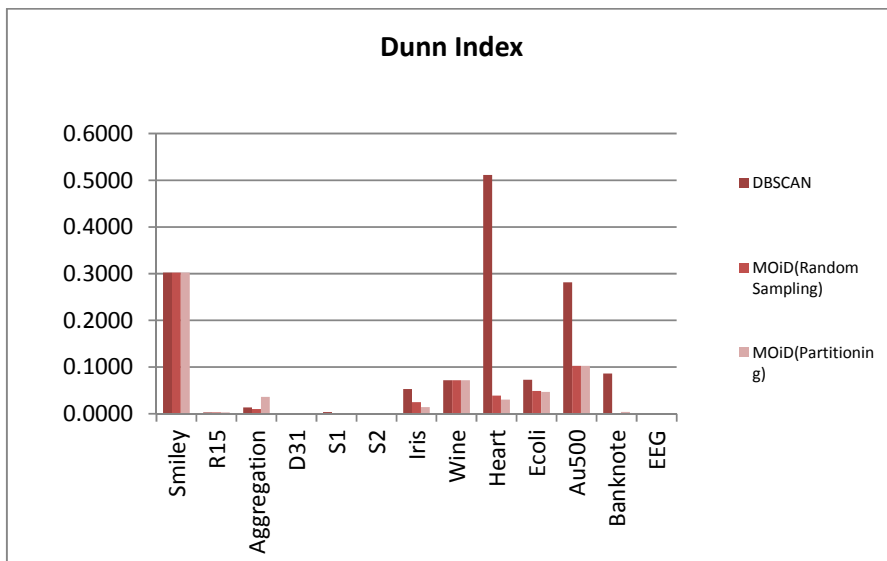


Fig. 11. Datasets Vs. Dunn Index of MOiD and DBSCAN

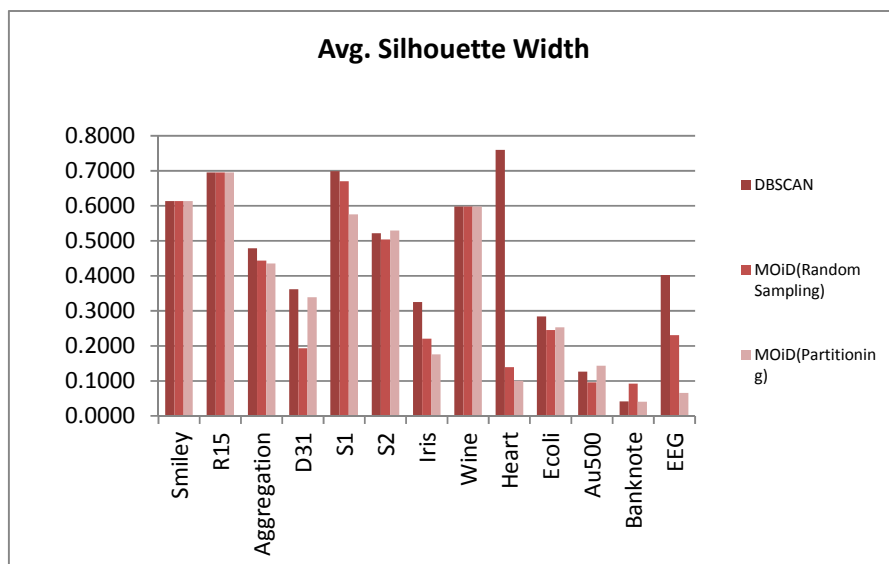


Fig. 12. Datasets Vs. Avg. Silhouette Width of MOiD and DBSCAN

The following tables show the overall statistics of the experiments for both kinds of datasets.

Table 2 : Performance Comparison of static DBSCAN and MOiD in terms of No. Of Region Query and Execution time

Increment Size	Total size of Datas et	No. of Region Query				Execution time (in Sec.)			
		Random Sampling		Partitioning		Random Sampling		Partitioning	
		DBS CAN	MOi D	DBS CAN	MOi D	DBSC AN	MOiD	DBSC AN	MOiD
		Smiley							
50	300	455	138	593	176	0.07	0.02	0.06	0.02
	350	505	146	647	208	0.06	0.03	0.11	0.03
	400	554	207	697	205	0.25	0.04	0.11	0.03
	450	604	163	747	206	0.31	0.03	0.14	0.03
	500	654	157	797	206	0.38	0.03	0.25	0.03
		R15							
50	400	807	759	812	238	0.1	0.09	0.08	0.03

	450	920	759	914	596	0.17	0.09	0.09	0.06
	500	1025	847	1017	541	0.2	0.19	0.18	0.06
	550	1119	827	1119	433	0.28	0.19	0.23	0.05
	600	1221	729	1218	294	0.3	0.09	0.3	0.02
	Aggregation								
50	588	1222	360	1209	209	0.12	0.02	0.14	0.03
	638	1324	400	1316	248	0.15	0.03	0.14	0.04
	688	1424	442	1420	305	0.17	0.03	0.21	0.05
	738	1522	481	1522	309	0.17	0.03	0.17	0.05
	788	1622	527	1623	154	0.21	0.03	0.19	0.02
	D31								
100	2700	5195	3932	4888	3437	2.55	1.25	2.04	1.25
	2800	5313	4030	5076	2290	2.49	1.13	2.23	1.13
	2900	5438	4128	5262	1333	3.05	1.19	2.42	1.19
	3000	5538	4227	5454	520	3.34	0.04	2.55	0.04
	3100	5645	4324	5644	1812	3.77	1.06	2.69	1.06
	S1								
200	4200	9148	1230	8345	1406	5.56	1.14	4.53	1.14
	4400	9350	1185	8745	1092	5.62	1.13	4.93	1.13
	4600	9551	1126	9148	2604	5.93	1.12	5.61	1.12
	4800	9749	1205	9546	1229	6.15	1.13	6.4	1.13
	5000	9947	1124	9947	1008	6.32	1.12	6.71	1.12
	S2								
200	4200	8588	1453	7762	2896	5.76	1.14	4.36	1.19
	4400	8796	1423	8199	2606	5.8	1.12	4.85	1.39
	4600	9003	1277	8580	1833	6.24	1.11	5.58	1.17
	4800	9207	1460	8951	2957	6.35	1.14	6.11	1.68
	5000	9407	1383	9406	3004	6.6	1.12	6.35	1.74
	IRIS								
10	110	210	77	207	81	0.02	0.01	0.03	0.01
	120	228	78	217	86	0.03	0.01	0.05	0.01
	130	252	91	245	125	0.05	0.01	0.05	0.02
	140	275	87	263	75	0.05	0.01	0.03	0.01
	150	294	101	291	112	0.05	0.02	0.05	0.01
	Wine								

10	138	284	36	281	81	0.05	0.01	0.05	0.01
	148	304	36	301	86	0.06	0.01	0.06	0.01
	158	323	38	321	125	0.07	0.01	0.07	0.01
	168	342	37	341	75	0.07	0.01	0.07	0.01
	178	361	39	361	112	0.07	0.01	0.07	0.01
Heart									
20	190	384	48	384	46	0.03	0.01	0.03	0.01
	210	424	47	424	27	0.04	0.01	0.03	0
	230	464	52	463	47	0.04	0.01	0.05	0.01
	250	503	48	504	50	0.06	0.01	0.05	0.01
	270	543	31	544	29	0.07	0	0.05	0
e-Coli									
20	256	461	156	409	110	0.03	0.01	0.04	0.02
	276	496	161	219	145	0.04	0.02	0.02	0.02
	296	540	163	223	167	0.06	0.02	0.02	0.02
	316	575	148	231	154	0.07	0.01	0.02	0.03
	336	609	181	171	146	0.07	0.02	0.01	0.03
AUS00									
50	300	590	115	591	112	0.3	0.2	0.28	0.03
	350	693	117	694	151	0.55	0.2	0.47	0.07
	400	796	161	793	211	0.59	0.2	0.67	0.11
	450	898	205	893	237	0.75	0.2	0.86	0.11
	500	998	238	998	292	1.14	0.2	1.03	0.1
Banknote									
100	972	1968	817	1960	490	0.67	0.08	0.95	0.07
	1072	2166	675	2165	525	0.81	0.08	0.96	0.06
	1172	2363	616	2364	487	1	0.07	1.03	0.05
	1272	2563	649	2564	723	1.36	0.06	1.33	0.07
	1372	2764	680	2762	493	1.47	0.07	1.53	0.05
EEG									
250	2543	5052	1099	5062	1117	4.97	0.13	4.87	0.16
	2793	5554	1271	5566	962	5.28	0.15	5.6	0.15
	3043	6072	1421	6080	952	6.35	0.15	6.76	0.15
	3293	6580	1200	6585	816	7.53	0.14	7.68	0.11
	3543	7102	1272	7102	1027	8.5	0.15	9.32	0.17

Table 3 : Performance Comparison of static DBSCAN and MOiD in terms of different validity measures

		Accuracy	Corrected Rand	Entropy	Dunn Index	Avg. Silhouette Width
Smiley						
MOiD	Random Sampling	100.00	1.0000	1.3074	0.3023	0.6134
	Partitioning	100.00	1.0000	1.3074	0.3023	0.6134
DBSCAN		100.00	1.0000	1.3074	0.3023	0.6134
R15						
MOiD	Random Sampling	92.75	0.9275	2.7696	0.0029	0.6953
	Partitioning	92.75	0.9275	2.7696	0.0029	0.6953
DBSCAN		92.75	0.9275	2.7696	0.0029	0.6953
Aggregation						
MOiD	Random Sampling	87.06	0.8706	1.6476	0.0103	0.4438
	Partitioning	99.84	0.9984	1.7025	0.0358	0.4356
DBSCAN		96.03	0.9603	1.7872	0.0134	0.4789
D31						
MOiD	Random Sampling	31.08	0.3108	3.1506	0.0001	0.1932
	Partitioning	47.91	0.4791	3.1319	0.0007	0.3388
DBSCAN		45.64	0.4564	3.2582	0.0007	0.3621
S1						
MOiD	Random Sampling	96.29	0.9629	2.7438	0.0003	0.6702
	Partitioning	87.34	0.8734	2.6380	0.0001	0.5756
DBSCAN		100.00	1.0000	2.7451	0.0034	0.6991
S2						
MOiD	Random Sampling	99.17	0.9917	2.7500	0.0012	0.5040
	Partitioning	94.57	0.9457	2.7519	0.0012	0.5297

DBSCAN		100.00	1.0000	2.7447	0.0012	0.5218
IRIS						
MOiD	Random Sampling	72.00	0.6211	1.4760	0.0245	0.2206
	Partitioning	62.67	0.4141	1.2233	0.0142	0.1760
DBSCAN		78.67	0.6841	1.4492	0.0530	0.3253
Wine						
MOiD	Random Sampling	60.45	0.2870	0.6999	0.0716	0.5982
	Partitioning	60.45	0.2870	0.6999	0.0716	0.5982
DBSCAN		60.45	0.2870	0.6999	0.0716	0.5982
Heart						
MOiD	Random Sampling	55.56	0.0048	0.1833	0.0388	0.1392
	Partitioning	57.04	0.0086	0.1166	0.0303	0.1006
DBSCAN		55.19	-0.0015	0.0244	0.5115	0.7596
e-Coli						
MOiD	Random Sampling	52.98	0.5244	1.2785	0.0490	0.2450
	Partitioning	52.98	0.4226	1.0415	0.0467	0.2533
DBSCAN		53.87	0.4190	1.0243	0.0726	0.2842
AU500						
MOiD	Random Sampling	51.80	0.0031	0.3815	0.1026	0.0961
	Partitioning	57.20	0.1687	0.6204	0.1026	0.1432
DBSCAN		55.60	0.0062	0.0737	0.2816	0.1266
Banknote						
MOiD	Random Sampling	71.94	0.4309	1.4584	0.0000	0.0919
	Partitioning	70.92	0.4412	1.4334	0.0040	0.0406
DBSCAN		84.62	0.5584	0.9320	0.0862	0.0412
EEG						
MOiD	Random Sampling	50.30	-0.0008	0.3334	0.0000	0.2304
	Partitioning	46.37	-0.0016	0.6681	0.0000	0.0658

DBSCAN	50.61	-0.0033	0.1812	0.0000	0.4024
---------------	-------	---------	--------	--------	--------

Table 2 displays the comparison results of static DBSCAN and our proposed algorithm, MOiD, in terms of number of region queries and the execution time. As shown in table 2, and in figure 4 and 5, the number of region queries performed by MOiD is very less than that of DBSCAN. The reason behind this is the size of the dataset which is to be clustered. In the static DBSCAN approach, whenever insertion is to be done, the whole updated dataset is clustered again, without taking into account the earlier cluster analysis. Hence, as the size of dataset is increased, the execution time is also increased with each increment. On the contrary, in MOiD, the increments are separately clustered and the resulting clusters are combined with the existing clusters. Thus the number of region queries and therefore the execution time is reduced at a great extent.

Table 3 displays the quality of the clusters in terms of rand index, entropy, Dunn index and silhouette width, for MOiD and DBSCAN.

5 Conclusion

The traditional clustering algorithms are only suitable for the static datasets. In the dynamic environment where the data is regularly updated, the clustering process becomes difficult and the results become unreliable too! This certainly decreases the efficiency and wastes the computing resources in order to form the clusters with all the data again. This research proposes a fundamentally different algorithm, MOiD, which works on multiple objects in contrast to the traditional incremental DBSCAN approach working on single object. MOiD, first, adds the data points in bulk by performing the clustering and then merges the clusters with the existing clusters. The experiment results are promising and much better than DBSCAN in terms of time and region queries. The other performance measures are also equally good or better in comparison with DBSCAN. In general, the overall performance of MOiD beats the DBSCAN. One thing which can be inferred is that if the clusters are well separated then MOiD performs the best over DBSCAN. However, the performance of MOiD may vary depending on the distribution of the new data to be added into the existing clusters, which is one of the characteristics of DBSCAN.

References

- [1] Jain, A. K. (2010) Data Clustering: 50 Years Beyond K-Means. Pattern Recognition Letters, 31, 651-666.
- [2] Han, J., Kamber M., (2001) Data Mining: Concepts and Techniques. Morgan Kaufman, Elsevier.
- [3] Berkhin, P. (2006) A survey of clustering data mining techniques. In Grouping multidimensional data. Springer, Berlin, Heidelberg.

- [4] Jain, A.K., Murty, M. N., Flynn, P. J. (1999) Data clustering: a review. *ACM Computing Surveys*, 31, 264-32.
- [5] Xu, R., Wunsch, D. (2005) Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16, 645-678.
- [6] Soni N., Ganatra, A. (2012) Categorization of several clustering algorithms from different perspective: a review. *International Journal of Advanced Research in Computer Science and Software Engineering*, 1, 1-6.
- [7] Jain, A. K., Dubes, R. C. (1988) Algorithms for clustering data. Prentice-Hall, Inc.
- [8] Zhou, S., Zhou, A., Jin, W., Fan, Y., Qian, W. (2000) FDBSCAN: a fast DBSCAN algorithm. *RUAN JIAN XUE BAO*, 11, 735-744.
- [9] Viswanath, P., Pinkesh, R. (2006) l-dbscan: A fast hybrid density based clustering method. *Proceedings of Int. Conf. On Pattern Recognition '06*, Hong Kong, August, pp. 912-915. IEEE.
- [10] Jian, L., Wei, Y., Bao-Ping, Y. (2009) Memory effect in DBSCAN algorithm. *Proceedings of 4th Int. Conf. on Computer Science & Education*, Nanning '09, China, 12th Nov., pp. 31-36. IEEE.
- [11] Peter, J. H., Antonysamy, A. (2010) An Optimised Density Based Clustering Algorithm. *International Journal of Computer Applications*, 6, 20-25.
- [12] Cassisi, C., Ferro, A., Giugno, R., Pigola, G., Pulvirenti, A. (2013) Enhancing density-based clustering: Parameter reduction and outlier detection. *Information Systems*, 38, 317-330.
- [13] Darong, H., Peng, W. (2012) Grid-based DBSCAN algorithm with referential parameters. *Physics Procedia*, 24, 1166-1170.
- [14] Fahim, A. M., Saake, G., Salem, A. M., Torkey, F. A., Ramadan, M. A. (2009) An enhanced density based spatial clustering of applications with noise. *Proceedings of DMIN '09*, Las Vegas, USA, pp. 517-523,. CSREA Press.
- [15] Fahim, A. M., Saake, G., Salem, A. M., Torkey, F. A., Ramadan, M. A. (2009) Dcbor: a density clustering based on outlier removal. *International Journal of Computer Science*, 4.
- [16] Borah, B., Bhattacharyya, D. K. (2008) DDSC: a density differentiated spatial clustering technique. *Journal of Computers*, 3, 72-79.

- [17] Liu, P., Zhou, D., Wu, N. (2007) VDBSCAN: varied density based spatial clustering of applications with noise. Proceedings of Int. Conf. on Service Systems and Service Management '07, Chengdu, China, 9-11 June, pp. 1-4. IEEE.
- [18] Birant, D., Kut, A. (2007) ST-DBSCAN: An algorithm for clustering spatial-temporal data. *Data & Knowledge Engineering*, 60, 208-221.
- [19] Biçici, E., Yuret, D. (2007) Locally scaled density based clustering. Proceedings of Int. Conf. on Adaptive and Natural Computing Algorithms '07, Warsaw, Poland, 11-14 April, pp. 739-748. Springer, Berlin, Heidelberg.
- [20] Song, Y. C., Meng, H. D., Wang, S. L., O'Grady, M., O'Hare, G. (2009) Dynamic and Incremental Clustering Based on Density Reachable. Proceedings of Fifth Int. Joint Conf. on INC, IMS and IDC '09, Seoul, 25-27 Aug., pp. 1307-1310. IEEE.
- [21] Davidson, I., Ravi, S. S., Ester, M. (2007) Efficient incremental constrained clustering. Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 240-249. ACM.
- [22] Fraley, C., Raftery, A., Wehrens, R. (2005) Incremental model-based clustering for large datasets with small clusters. *Journal of Computational and Graphical Statistics*, 14, 1-18.
- [23] Can, F. (1993) Incremental clustering for dynamic information processing. *ACM Transactions on Information Systems*, 11, 143-164.
- [24] Hsu, C. C., Huang, Y. P. (2008) Incremental clustering of mixed data based on distance hierarchy. *Expert Systems with Applications*, 35, 1177-1185.
- [25] Inekave, S., Last, M., Maimon, O. (2007) Incremental clustering of mobile objects. Proceedings of 23rd International Conference on Data Engineering Workshop '07, Istanbul, 17-20 April, pp. 585-592. IEEE.
- [26] Patra, B. K., Nandi, S., & Viswanath, P. (2011) A distance based clustering method for arbitrary shaped clusters in large datasets. *Pattern Recognition*, 44, 2862-2870.
- [27] Charikar, M., Chekuri, C., Feder, T., Motwani, R. (2004) Incremental clustering and dynamic information retrieval. *SIAM Journal on Computing*, 33, 1417-1440.
- [28] Chen, C. Y., Hwang, S. C., Oyang, Y. J. (2002) An incremental hierarchical data clustering algorithm based on gravity theory. In *Advances in knowledge discovery and data mining* (pp. 237-250). Springer, Berlin, Heidelberg.
- [29] Ester, M., Kriegel, H. P., Sander, J., Xu, X. (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining, Portland, August, pp. 226-231.

- [30] Ester, M., Kriegel, H. P., Sander, J., Wimmer, M., Xu, X. (1998) Incremental clustering for mining in a data warehousing environment. Proceedings of 24th International Conference on Very Large Data Bases, August, pp. 323–333, 1998.
- [31] Veenman, C. J., Reinders, M. J., Backer, E. (2002) A maximum variance cluster algorithm. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24, 1273-1280.
- [32] Gionis, A., Mannila, H., Tsaparas, P. (2007) Clustering aggregation. ACM Transactions on Knowledge Discovery from Data, 1, 1-30.
- [33] Fränti, P., Virtajoki, O. (2006) Iterative shrinking method for clustering problems. Pattern Recognition, 39, 761-775.
- [34] C. Hennig, “fpc: Flexible Procedures for Clustering”, R package version 2.0, Available at: <http://CRAN.R-project.org/package=fpc>, 2010.
- [35] Manual for the DBSCAN algorithm implemented on FPC, Available at http://bm2.genes.nig.ac.jp/RGM2/R_current/library/fpc/man/dbscan.html
- [36] Rand, W. M. (1971) Objective criteria for the evaluation of clustering methods. Journal of the American Statistical association, 66, 846-850.
- [37] Yeung, K. Y., Ruzzo, W. L. (2001) Details of the adjusted Rand index and clustering algorithms, supplement to the paper “An empirical study on principal component analysis for clustering gene expression data”, Bioinformatics, 17, 763-774.
- [38] Dunn, J. C. (1973) A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters, Journal of Cybernetics, 3, 32–57.
- [39] Rousseeuw, P. J. (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20, 53-65, 1987.
- [40] <https://archive.ics.uci.edu/ml/datasets.html>

Wavelet based OFDM with ICI Self-Cancellation in Underwater Acoustic Communications

Naresh Kumar, *Member, IEEE* and B. S. Sohi, *Sr. Member, IEEE*

Abstract—There are many research challenges in underwater acoustic communication environment such as large delay spread, ocean waves, motion of transmitter/receiver, Doppler spread etc. OFDM has potential to combat with many such problems, but it is also deteriorated by Inter Carrier Interference and high peak to average power ratio. Conventional OFDM is spectral inefficiency as it uses cyclic prefixing which consumes approximately 20% of available bandwidth. ICI self cancellation technique performs better for ICI problems. As it transmits redundant data on adjacent subcarriers which makes some subcarriers idle, hence, ICI is reduced at the cost of bandwidth. In this paper, a Wavelet based OFDM with ICI cancellations is proposed to counter the problem of ICI. Use of Wavelets reduces the need for cyclic prefixing thereby making it more spectral efficient and wavelets also help in maintaining orthogonality between subcarriers which further improves its ICI performance. Simulation results show that proposed technique performs better in terms of bit error rate (BER) as compared to conventional OFDM.

Index Terms — OFDM, Wavelets, BER, Self-Cancellations, ICI.

I. INTRODUCTION

Underwater Acoustic Sensor Networks can be used for exploring undersea resources and for gathering scientific data. Many applications including pollution monitoring, distributed tactical surveillance, ocean sampling networks, offshore exploration etc can be made possible by deploying underwater sensor networks along with Autonomous Underwater Vehicles (AUVs) [1-5]. Major challenge in implementation of underwater acoustic sensor networks lies in physical layer design. Underwater acoustic channel creates major obstacle in terms of delay spread, Doppler spread, and multipath propagation. Orthogonal Frequency Division Multiplexing (OFDM) is used in underwater acoustic communication as it is capable of combating Inter Symbol Interference (ISI) generated by large delay spread [6-8]. OFDM converts frequency selective fading channel into multiple orthogonal sub carriers into flat fading. If the orthogonality between sub carriers is lost because of one problem or other then there arises the problem of ICI. Possible

causes of ICI are Doppler shift because of movement of transmitter and receiver, large scale fading introduced by underwater channel. In an FFT based OFDM system cyclic prefix (CP) is introduced after each symbol frame to reduce the effect of ISI. This CP inserted between subcarriers consumes approximately 20 % of bandwidth and makes OFDM system spectrally inefficient. Other problems of FFT based OFDM are ICI and high PAPR. These problems greatly affect the performance of the system. Here, in this present paper ICI problem is considered and techniques which counter this problem are discussed. In the literature, various ICI countermeasures are presented by researchers [9-15]. ICI self cancellation techniques are simple to implement. In ICI self cancellation technique adjacent-mapping-based ICI cancellation is proposed as in [15] making some subcarriers as idle and ICI is reduced. Many techniques have proposed to improve spectral efficiency of ICI self cancellation [16-18] but this improvement effects the ICI performance. In this paper Wavelet based OFDM is integrated with ICI self cancellation for better performance of ICI cancellation.

II. OFDM MODULATION

A. OFDM Transmitter

A conventional FFT based OFDM system is described in Fig.1. OFDM system transmits the data by modulating these into different sub carriers. In an OFDM system, If N number of subcarriers is there then these subcarriers will be transmitted and written as $Nu + 1$. It will lie on central spectrum and the subcarriers at the edges will form the guard bands. These subcarriers are modulated using a data symbol $X_{a,n}$ where 'a' is the number of OFDM symbol and 'n' is subcarrier number. Inverse Fast Fourier Transform (IFFT) of size N is then applied. The subcarriers in the guard band are not utilized in order to keep the size of the transmit signal less than the bandwidth size of $1/T$. T is sampling time of OFDM signal. A guard interval helps combat the inter symbol interference in a multipath fading channel environment. The resultant signal at the output of the transmitter can be written as [19]

$$s(t) = \sum_{a=-\infty}^{\infty} \sum_{n=-\frac{Nu}{2}}^{\frac{Nu}{2}} X_{a,n} \psi_{a,n}^* \otimes g_T(t) \quad (1)$$

Where, \otimes represents the convolution, impulse response of the analog filter used in the transmission is given by $g_T(t)$ and $\psi_{a,n}(t)$ stands for the subcarrier pulse.

Sub carrier pulse $\psi_{a,n}(t)$ is given below:

Corresponding author is Naresh Kumar.

Naresh Kumar is working as Assistant Professor in Electronics & Communication Engineering branch of University Institute of Engineering & Technology (UIET), Panjab University Chandigarh, India. His areas of interest includes wireless communication, Radio over Fibre, Underwater Communication etc.

B. S. Sohi is presently working as Pro Vice-Chancellor at Chandigarh University, Gharuan, Punjab, India. He has also worked as Director, University Institute of Engineering & Technology (UIET), Panjab University, Chandigarh, India. His areas of interest includes wireless communication, cross layer design, wireless sensor networks, embedded systems, antenna design, underwater acoustic communication etc.

$$\psi_{a,n}(t) = \begin{cases} e^{j2\pi \frac{n}{T_u}(t-\Delta-aT_s)} & aT_s \leq t \leq (a+1)T_s \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

Where, T_s is total time including guard time, Δ is the length of the guard interval. So, $T_s = T_u + \Delta$, OFDM subcarrier spacing will be equal to $1/T_u$.

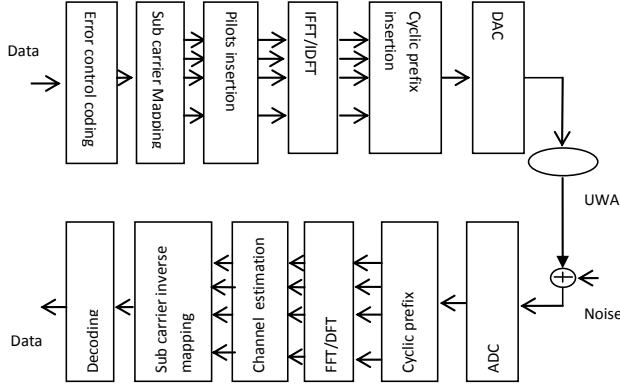


Fig.1 OFDM system for Underwater Acoustic Communication.

B. Underwater Acoustic Channel with noise

The signal $s(t)$ when transmitted through a linear time – invariant multipath channel [20] with additive Gaussian Noise results in a received signal given by:

$$r(t) = \sum_{p=0}^{N_p-1} h_p s(t - \tau_p) + w(t) \quad (3)$$

Where N_p = Number of propagation paths,
 h_p = path gain, τ_p = path delay of the p^{th} path, and $w(t)$ is the additive Gaussian Noise with power spectral density $\frac{N_0}{2}$.

C. OFDM receiver

At the OFDM receiver, by assuming that the guard interval longer than the channel delay spread, when synchronization is perfect, and then the n^{th} subcarrier output during the k^{th} OFDM symbol can be described as [21]

$$Y_{k,n} = X_{k,n} \cdot G_T(n) G_R(n) + n_{k,n}, \quad -\frac{N_u}{2} \leq n \leq \frac{N_u}{2} \quad (4)$$

Where, $n_{k,n}$ represents the additive white Gaussian noise, $G_T(n)$ stands for the frequency response of the analog transmission filter and $G_R(n)$ denotes the frequency response of the receiver filter at the n^{th} subcarrier frequency $f_n = \frac{n}{T_u}$, channel response in the frequency domain is denoted as $H_{a,n}$ and can be mathematically explained as [21].

$$H_{k,n} = \sum_{r=1}^R h_n(kT_s) \cdot e^{-j2\pi \frac{nT_r}{NT}} \quad (5)$$

The channels r^{th} path gain during the k^{th} number of the OFDM symbol is represented by in Equation (5). If our transmission subcarriers lie in the flat region of the analog

filters both at the transmitter and the receiver then we can rewrite the Equation (4) as [11]

$$Y_{k,n} = X_{k,n} \cdot H_{k,n} + n_{k,n} \quad (6)$$

Which mean we have assumed $G_T(n)$ and $G_R(n)$ to be equal to one in their flat region. Another way to eliminate $G_T(n)$ and $G_R(n)$ from Equation (4) will be to use the a priori knowledge of the transmitter and receiver filters.

III. WAVELET BASED OFDM MODULATION

Wavelet based OFDM modulation is an alternate technique to conventional FFT based OFDM. In terrestrial networks, wavelet based OFDM has shown same advantages to that of conventional FFT based OFDM and moreover, PAPR reduction and combat frequency timing offset are addition benefits [22]. Wavelet based OFDM fulfills the condition for orthogonality. This system shows perfect reconstruction when tested with orthogonal filters of Quadrature mirror filter bank. Discrete Wavelet Transform (DWT) has large power spectral density in comparison with conventional OFDM. This large power spectral density is because it produces well contained main lobe whereas its narrow side lobes are having reduced out-of-band emissions. Wavelets are also having multi-resolution capability in which signal is well located both in time and frequency domain [23]. Wavelets are used in place of DFT/FFT as an alternative transform.

A. Wavelet back ground

Wavelet Transform can be used to decompose a continuous time signal. When a signal is passed through wavelets it produces signal into different scales and different times [24]. Wavelets are having multi-resolution capabilities. Continuous Wavelet Transform can be written as:

$$\begin{aligned} CWT x(\tau, b) &= \langle x(t), \Psi_{b,\tau} \rangle \\ &= \frac{1}{\sqrt{|b|}} \int_{-\infty}^{\infty} x(t) \Psi^* \left(\frac{t-\tau}{b} \right) dt \end{aligned} \quad (7)$$

Where ' Ψ ' represents the mother wavelet, parameter ' τ ' is the translation which corresponds to the time information present in the signal and ' b ' is the scaling parameter which corresponds to the frequency information contained in the signal and ' $*$ ' represents the complex conjugate [25]. Real Signals (zero complex value) doesn't need complex conjugate. So, for real signal, complex conjugate of ' Ψ ' is not required.

A mother wavelet is designed in such a way that it can be inverted to retrieve the original transmitted signal. For continuous wavelet transform there exists no viable inverse in practice due to the redundancy in the information which will require extensive analytical calculation, but in theory the inverse for this transform can be written as [26].

$$x(t) = \frac{1}{c_\psi} \int_a \int_{\tau} \frac{1}{a^2} x(a, \tau) \Psi \frac{(t-\tau)}{a} dt da \quad (8)$$

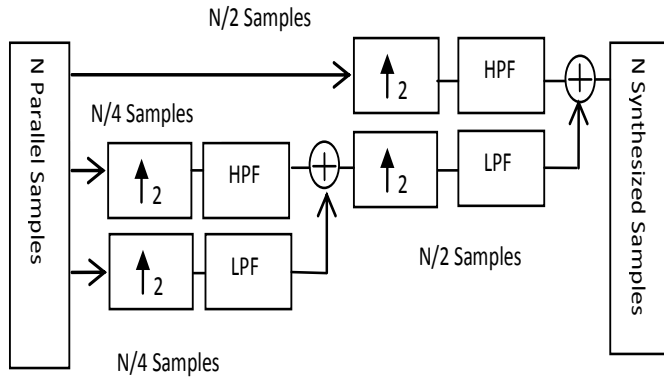


Fig. 2. Two-Level Wavelet Synthesis

Now to circumvent the above mentioned data redundancy issue we discretize the scaling and in addition the translation variables. Now using the equation (7) as a reference if we discretize the translation parameter ' τ ' by $2^j k$, and scaling parameter ' a ' by 2^j , then we can rewrite equation (7) as given below:

$$\begin{aligned} DWT x(j, k) &= \langle x(t), \Psi_{j,k} \rangle \\ &= 2^{-\frac{j}{2}} \sum_{n=-\infty}^{\infty} x(n) \Psi(2^{-j}n - k) \end{aligned} \quad (9)$$

Equation (9) is discrete wavelet transform of signal $x(t)$. This transform can also be referred to as a form of sub-band coding because in order to analyse a signal it has to pass through a string of filter banks [25]. Every such signal passes through a high pass filter and a low pass filter.

B. DWT scheme and Reconstruction of Signal

The process involved in implementing the wavelet based OFDM is similar to the process involved in conventional FFT-OFDM which is as follows:

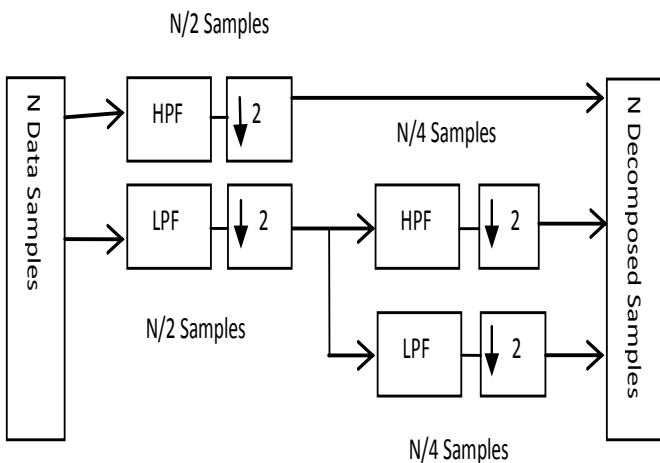


Fig. 3 Two-Level Wavelet Decomposition

Wavelet transform based OFDM systems consist of perfect reconstruction quadrature mirror filter bank that employ half

band low pass filter (LPF) whose impulse response can be written as ' h ' and half band high pass filter (HPF) whose impulse response can be written as ' g '. These two filters HPF & LPF convolve with the input signal $x[n]$. Two sequential parallel data streams are used in wavelet transform multiplexing at every scale and translation i.e. $x_{low}(n)$ and $x_{high}(n)$ which is described in Fig. 2. This data samples are then up sampled by a factor of 2 and passed through the LPF and HPF of QMF bank. The input signal to each filter convolves with the impulse response of the filter to yield $x_{low}[n] = h[n] * x[n]$ for the low pass filter, which gives us our approximate coefficient and $x_{high}[n] = g[n] * x[n]$ for the HPF. LPF produces the approximation coefficient at the filter output and HPF produces detailed coefficient. Anti imaging filter is required in order to filter out the image frequencies that are produced during the up sampling operation within each channel. These filtered streams are summed and constitute a wavelet symbol. This process is synthesis process also known as inverse DWT-IDWT process. This synthesized data is passed through the channel in the presence of AWGN.

When the signal is received on the receiver side it is then again passed through the Quadrature mirror filter bank which consists of a pair of conjugate LPF $h^*(-n)$ and conjugate HPF $g^*(-n)$ on the receiver end. Two such pairs constitute a two channel QMF bank as shown in Fig. 3. The received signal is first decomposed into its respective detailed and approximate coefficients and then down sampled by a factor of 2 which is also apparent from the Fig. 3. This process continues until the N parallel streams of data are recovered. The recovered data is then converted into a serial stream using a parallel to serial converter and then demodulated using a suitable scheme. In this study we have used 16-QAM modulation.

The HPF and LPF of the QMF bank are expressed mathematically as:

$$h(n) = (-1)^n g(L + 1 - n) \quad (10)$$

Where, L is the sequence of length of $g(n)$. Because of the shift and translation of the wavelet transformed signals, each of the composite symbols is rather delayed by a factor ' α ' according to the z -transform relation ($X(z) = \sum_n x(n)z^{-n}$, where $z^{-\alpha} = e^{-j\alpha\omega}$) which requires adjacent matched filters to perfectly reconstruct the signals. This perfect reconstruction can only hold if the matched filters respect the following.

$$h(z)h^*(z) + g(z)g^*(z) = 2z^{-\alpha} \quad (11)$$

$$h(z)h^*(-z) + g(z)g^*(-z) = 0 \quad (12)$$

From above processes, it is clear that cyclic prefixing which is required in conventional OFDM is not required in Wavelet based OFDM whereas symmetric extension may be used.

IV. ICI SELF-CANCELLATION SCHEME

In [15], the ICI self cancellation scheme based on a data allocation of $(X(k), X(K + 1) = -X(k))$, $k = 0, 2, \dots, N - 2$, has been proposed to deal with the intercarrier interference. The received signal $Y(k)$ is determined by the difference between the adjacent subcarriers.

Assume the transmitted symbols are constrained, then, the

received signal within k^{th} subcarrier becomes as

$$Y'(k) = \sum_{l=0}^{N-1} X(l)S(l-k) + n(k) \quad (13)$$

The demodulation is designed to work in such a way that each signal at the $k+1^{\text{th}}$ subcarrier (where k denotes even number) is multiplied by “-1” and then summed with the one at the k^{th} subcarrier. Then the resultant data sequence is used for making symbol decision. It can be given as:

$$Y''(k) = Y'(k) - Y'(k+1) \quad (14)$$

The sequence $S(l-k)$ is defined as the ICI coefficient between l^{th} and k^{th} subcarriers.

$$S''(l-k) = -S(l-k-1) + 2S(l-k) - S(l+1-k) \quad (15)$$

According to the definition of the Carrier to Interference Ratio (CIR), can be represented as below:

$$\text{CIR} = \frac{|-S(-1)+2S(0)-S(1)|^2}{\sum_{l=2}^{N-2} \text{even} |-S(l-1)+2S(l)-S(l+1)|^2} \quad (16)$$

C. Proposed Scheme

There are advanced transforms techniques like Wavelet Transform and others which can be used in the place of Fourier transform [28]. In the proposed design we are using inverse discrete wavelet transform instead of inverse discrete Fourier transform at transmitter and discrete wavelet transform at the place of discrete Fourier transform at receiver.

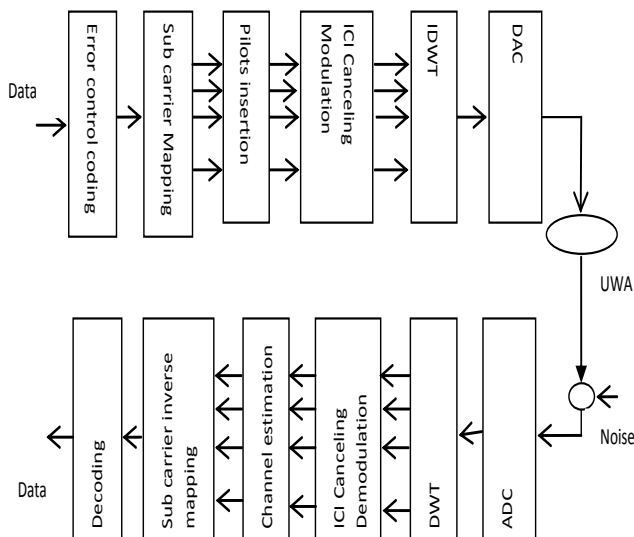


Fig. 4. Wavelet based OFDM with ICI self cancellation for underwater acoustic communication.

This Wavelet based OFDM is integrated with ICI Self Cancellation. Wavelet transform is a tool for the analysis of signal in time and frequency domain. Here in this analysis mechanism input signal is decomposed into different

frequency components [29]. Wavelet gives better orthogonality and have localization in both time and frequency domain [30]. Wavelet based multicarrier communication has been recognized as good option in different works [27, 31-34]. Wavelet based system gives better BER performance as compared to conventional system in terrestrial networks. As in the proposed design for the self cancellation modulation, here also we will use ICI self cancellation modulation [15,35] in the beginning. During that coding the same data on a particular subcarrier will be coded on the adjacent subcarriers. This method is called adjacent symbol repetition method. With the help of this method the bandwidth required is more. After doing this coding N – point IDWT is performed, where N is the number of subcarriers required. Then in the conventional system cyclic prefixing is done but in the case of wavelet transform there is no requirement of cyclic prefixing which is a potential advantage of this scheme. Then for the transmission the data is converted form parallel to serial form and digital to analog conversion is also performed. This data is passed through underwater acoustic channel in the presence of AWGN noise.

Then at the receiver side, received data having a frequency offset of $\Delta F t$ is processed as per Fig. 4, in this data we are having the presence of noise as well, which is Gaussian in nature. On the received data, first of all, analog to digital conversion is carried out to further process this digital data. In the convention system, where removal of cyclic prefixing was done but here in the case of wavelet transform no cyclic prefixing is being used. Wavelet based OFDM system is 20% more bandwidth efficient as compared to conventional system because carrier prefixing not required [29]. So after ADC N -point discrete wavelet transform is performed and then the ICI self cancellation demodulation is done and required data is obtained which is tested for bit error rate.

V. SIMULATION RESULTS

A For the purpose of simulation MATLAB is used and BER performance curves are obtained for different values of frequency offset. Different values of signal to noise ratio are taken for which different BER values are plotted. During simulation total 64 subcarriers are used, 12 subcarriers are used as pilots. In FFT based OFDM 20% cyclic prefixing is used whereas in Wavelet based OFDM no cyclic prefixing is used. 16 QAM modulations are being used. Then IDFT on conventional and IDWT in wavelet based system is performed on the data at the transmitter side. For the conventional system cyclic prefixing of 16 bits is used but in the case of wavelet based system no cyclic prefixing is being done. Then ICI Self cancellation modulation is performed as [15]. The values of SNR taken are 0 dB to 16 dB at the gap of 2 dB. UWA channel along with AWGN channel is used for the purpose of transmission of the data from the transmitter to the receiver. It is assumed that Doppler spread present is equal at all the paths of underwater channel and Carrier Frequency Offset (CFO) of .02, .05 and .08 are considered in the present study. At the receiver for the conventional system, first of all, the cyclic prefixing is removed before ICI Canceling demodulation

whereas in proposed system DWT is used before ICI Canceling demodulation. We have used db2 wavelet in the present study. Fig. 5, 6 and 7 show results of 16-QAM modulation and db2 wavelet at carrier frequency offsets of .02, .05, .08 respectively. From results it is observed that wavelet based self cancellation coded system is giving the better performance as compared to the conventional system. Performance of Wavelet based OFDM without using ICI Self Cancellation modulation is better than conventional OFDM without using ICI Self Cancellation method. It is also noteworthy that with ICI Self Cancellation modulation with conventional OFDM performs better than wavelet based OFDM without ICI Self Cancellation modulation as observed from Fig. 5, 6 and 7 at CFO of 0.02, 0.05 and 0.08. This proves ICI Self Cancellation method is better in handling ICI. Results also clearly indicate that Wavelet based OFDM with ICI Self Cancellation outperform conventional FFT based OFDM with ICI Self Cancellation method.

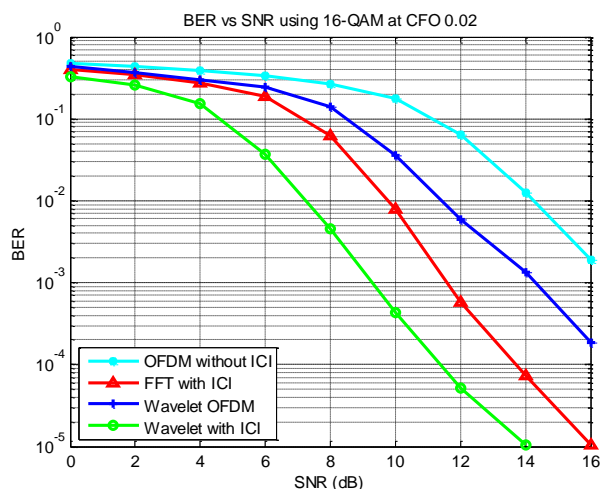


Fig. 5. BER vs SNR performance of conventional and proposed wavelet based system at carrier frequency offset of .02.

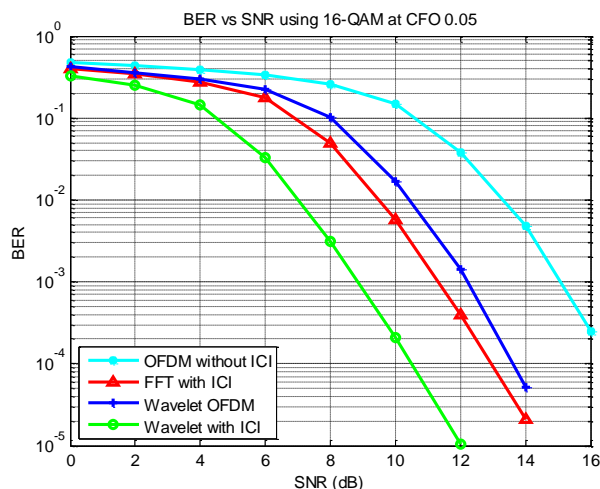


Fig. 6. BER vs SNR performance of conventional and proposed wavelet based system at carrier frequency offset of .05 using 16-QAM.

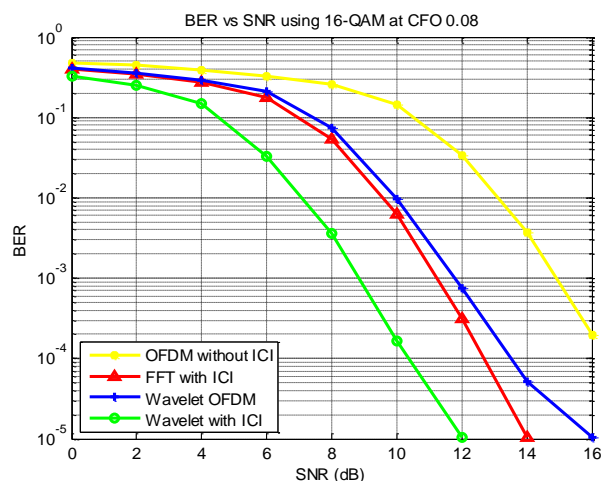


Fig. 7. BER vs SNR performance of conventional and proposed wavelet based system at carrier frequency offset of .08 using 16-QAM modulations.

VI. CONCLUSION

In the proposed design, wavelet based OFDM is used with ICI self cancellation. As the use of wavelets, orthogonality between subcarriers remains better; this results in better performance of the system for ICI cancellation. Results show that Wavelet based OFDM with ICI Self Cancellation modulation is better than Fourier based OFDM system with ICI Self Cancellation. It can be concluded that the proposed system is better as compared to the conventional system in terms of ICI cancellation performance. Secondly, cyclic prefixing is not used in proposed design which shows proposed system can perform better as compared to conventional system in terms of spectral efficiency. A hybrid technique can further improve the performance of ICI cancellation keeping the system bandwidth efficient.

VII. ACKNOWLEDGEMENT

The authors sincerely thank to Prof. Renu Vig, Director UIET for providing lab facilities to complete this task.

REFERENCES

- [1] I. F. Akyildiz, D. Pompili, and T. Melodia, "Underwater acoustic sensor networks: Research challenges," *Ad Hoc Networks*, Vol. 3(3), pp. 257–279, 2005.
- [2] I. F. Akyildiz, D. Pompili, and T. Melodia, "State-of-the-Art in Protocol Research for Underwater Acoustic Sensor Networks", in *Proceedings of ACM WUWNet'06*, 2006.
- [3] J.-H. Cui, J. Kong, M. Gerla, and S. Zhou, "Challenges: Building scalable mobile underwater wireless sensor networks for aquatic applications," *IEEE Network*, Special Issue on Wireless Sensor Networking, pp. 12–18, 2006.
- [4] J. Heidemann, W. Ye, J. Wills, A. Syed, and Y. Li, "Research challenges and applications for underwater sensor networking," in *Proceedings of the IEEE Wireless Communications and Networking Conference*, 2011.
- [5] J. Partan, J. Kurose, and B. N. Levine, "A survey of practical issues in underwater networks," in *Proceedings of ACM WUWNet'06*, 2006.
- [6] J. Proakis, E. M. Sozer, J. A. Rice and M. Stojanovic, "Shallow Water Acoustic Networks", *IEEE Communications Magazine*, Vol. 39 (11), pp.114-119, Nov 2001.

- [7] S. Climent, A. Sanchez, JV Capella, N. Meratnia, JJ. Serrano, "Underwater Acoustic Wireless Sensor Networks: Advances and Future Trends in Physical, MAC and Routing Layers", *Sensors* (Basel, Switzerland); Vol. 14(1), pp 795-833, 2014.
- [8] L. Yin, B. Chen, P. Cai, "Implementation and design of underwater acoustic speech communication system based on OFDM technology", *AASRI Procedia*, Vol. 1, pp. 46-51, 2013.
- [9] M. Wen, X. Cheng, X. Cheng, L. Yang, and B. Jiao, "Effective intercarrier interference reduction techniques for OFDM underwater acoustic communications," in *Proc. IEEE Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, USA, pp. 93-97, Nov. 2013.
- [10] X. Cheng, M. Wen, X. Cheng, D. Duan, and L. Yang, "Effective mirror-mapping-based intercarrier interference cancellation for OFDM underwater acoustic communications," *Ad Hoc Networks*, pp. 5-16, Vol. 34, Nov. 2015.
- [11] X. Cheng, M. Wen, X. Cheng, L. Yang, and Z. Xu, "Effective Self Cancellation of Intercarrier Interference for OFDM Underwater Acoustic Communications," in *Proc. ACM International Workshop on Underwater Networks (WUWNet13)*, Kaohsiung, Taiwan, Nov. 2013.
- [12] M. Wen, X. Cheng, X. Wei, B. Ai, and B. Jiao, "A novel effective ICI selfcancellation method," in *Proc. IEEE GLOBECOM*, Houston, TX, USA, Dec. 2011, pp. 1-5.
- [13] M. Wen, X. Cheng, L. Yang, and B. Jiao, "Two-path transmission framework for ICI reduction in OFDM systems," in *Proc. IEEE GLOBECOM*, Atlanta, GA, USA, Dec. 2013, pp. 3716-3723.
- [14] X. Cheng, Q. Yao, M. Wen, C. X. Wang, L. Song, and B. Jiao, "Wideband channel modeling and ICI cancellation for vehicle-to-vehicle communication systems," *IEEE J. Sel. Areas in Commun.*, vol. 31, no. 9, pp. 434-448, Sept. 2013.
- [15] Y. Zhao, S-G. Haggman, "Intercarrier interference self-cancellation scheme for OFDM mobile communication systems," *IEEE Trans. on Commun.*, vol. 49, no. 7, pp. 1185-1191, Jul. 2001.
- [16] A. Kumar and R. Pandey, "A spectrum efficient intercarrier interference reduction scheme for orthogonal frequency division multiplexing system in low signal to noise ratio environment" *IETE journal of research*, vol. 59, no. 1, January-February 2013.
- [17] A. Kumar and R. Pandey, "A bandwidth efficient method of cancellation of ICI in OFDM systems" *International journals of electronics and communication*, vol. 63, no. 7, pp. 569-575, July 2009.
- [18] M. Wen, Y. Li, X. Cheng and L. Yang, "Index modulated OFDM with ICI self-cancellation in underwater acoustic communications," *48th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, pp. 338-342, Nov. 2014.
- [19] Y. Baoguo, K.B. Letaief, R.S. Cheng, Z. Cao, "Channel estimation for OFDM transmission in multipath fading channels based on parametric channel modeling", *IEEE Transactions on Communications*, Vol. 49, No. 3, 2001, pp. 467- 479.
- [20] K. Davidson, "MACE10- Mobile acoustic communication experiment 2010-Quick-LookCruise Report", *Tech. Report*, Aug. 2010.
- [21] Ye Li, G.L. Stuber, *Orthogonal Frequency Division Multiplexing*, Springer science + business media inc., 2006.
- [22] V. Kumbasar and O. Kucur, "Better wavelet packet tree structures for PAPR reduction in WOFDM systems," *Digital Signal Processing*, 2008, Vol. 18, No. 6, pp. 885- 891.
- [23] M. Weeks, *Digital Signal Processing Using MATLAB and Wavelets*, Georgia State University, 2007.
- [24] M. K. Lakshmanan, and H. Nikookar, *A Review of Wavelets for Digital Wireless Communication*, *Wireless Personal Communications*, Vol.37, (2006), pp. 387-420.
- [25] I. Daubechies, *Ten lectures on wavelets*, Society for Industrial Mathematics, Vol. 61, 1992.
- [26] V. Kumbasar, et al., "Optimization of wavelet based OFDM for multipath powerline channel by genetic algorithm," *Wireless Communications and Mobile Computing*, Vol. 9, No. 9, pp. 1243-1250, 2009.
- [27] R. Asif, R.A. Abd-Alhameed, O.O.Anoh and Y.A.S. Dama, "Performance Evaluation of DWT-FDM and FFT-OFDM for Multicarrier Communications Systems using Time Domain Zero Forcing Equalization", *International Journal of Computer Applications*, Vol. 51 (4), August 2012.
- [28] P. Moose, "A technique for orthogonal frequency division multiplexing frequency offset correction", *IEEE transaction on communication*, vol. 42, no. 10, October 1994.
- [29] L. Madan Kumar, N. Homayoun, "A review of wavelets for digital wireless communication", *Wireless personal communications*, Kluwer academic publishers- Plenum publishers, vol. 37, no. 3-4, pp. 387-420, May 2006.
- [30] Broughton SA, Bryan K. *Discrete Fourier analysis and wavelets*. New Jersey, John Wiley, 2009.
- [31] M. K. Gupta and S. Tiwari, "Performance evaluation of conventional and wavelet based OFDM system", *international journal on electronics and communication*, vol. 67, no. 4, pp. 348-354, April 2013.
- [32] B. G. Negash and H. Nikookar, "Wavelet based multicarrier transmission over multipath wireless channels" *Electronic letters*, vol. 36, no.21, pp. 1787-1788. October 2000.
- [33] N. Kumar, G. Kaur, B.S. Sohi, " New modified channel estimation technique for ICI cancellation in wavelet based OFDM", *Far East Journal of Electronics and Communications* vol. 15 no.1, pp. 1-12, September 2015.
- [34] Anuradha, N. Kumar, " BER analysis of conventional and wavelet based OFDM in LTE using different modulation techniques" in *Proceedings of Engineering and Computational Sciences (RAECS)*,UIET Panjab University Chandigarh, India, IEEE, pp.1-4, March 2014.
- [35] N. Kumar, G. Kaur, B. S. Sohi, " Comparative analysis of various Inter-Carrier Interference methods", *I.J.Microwave and technologies, MECS press*, vol. 3, no.5, pp. 18-32, 2015.

A Method for Mining Social Media to Discovering Influential Users

Hosniyeh S. Arian¹ , Omid R. B. Speily²

1 Department of Computer Engineering and Information Technology, Islamic Azad University
Qazvin, Iran.

2 Department of Computer Engineering and Information Technology , Urmia University of
Technology
Urmia, Iran

Corresponding author : Hosniyeh S. Arian¹

ABSTRACT

Influential users who diffuse information and their followers have interest to this information finally they can maximize diffusion in social networks. Influential users have different influence in diversity domain specificity for instance user may have strong influence in a special topic and another topics have weak influence. So a proposed method presented for identifying influential users based on domain specificity in this paper. This method identified influential users based on domain specificity that features of user's profile and user's actions (e.g. retweet) that influence on diffusion determined by "multiple regression" and user's contents categorized based on keywords by "TF-IDF" and finally influential users identified by "Tree Regression" based on domain specificity in this paper. The detail of this method discussed the following of paper. In order to evaluate the proposed method on Twitter offer application program interface. 420 users selected randomly, they follow their friends, join to different groups, and generated diversity tweets on Twitter. The main feature, which distinguishes this method from the previously reported methods, is in two key respective. First previous studies have quantified influence in terms of network metrics for instance number of retweet or page rank, our proposed method measured influence in terms of the size Tree Regression. Second the focuses of previous studies were based on the structural of diffusion and feature of content but Influential users have different influence in diversity domain specificity so in our proposed method focused on this feature. Results showed that accuracy of proposed method is 0.69.

Keywords: Social networks, Categorized, Influence, Content, Diffusion, Domain specificity.

I. INTRODUCTION

Social networks generated a suitable platform for receive information between users. Users share their interesting and influence the ideas each other. so social networks are a golden opportunity for marketer and advertisers. Marketers

¹ Hosniyeh.safiarian@gmail.com , 09358857179, 021-88219140

can focus on the behavior of customers on social networks and acquirer information about their interesting and segment their customers accurately. Also the effectiveness of advertisement distribution depends on understanding information customer's segmentation. The advertisers have lack of appropriate advertising mechanism, they improve advertising mechanism with social networks.

The diffusion of information is fast and users can receive news in the short of time on social networks. So this media is an important tool for electronic word of mouth marketing. Electronic Word-of-Mouth (eWOM) is any positive or negative statement made by potential, actual, or former customers about a product or company, which is made available to a multitude of people and institutions via the Internet [3], and is a popular research topic in IS and marketing research [4]. Electronic word-of-mouth (eWOM) can reach large numbers of consumers [5]. And influence attitudes [6], product judgments [7], and sales [8] for instance 83 percent of Internet shoppers reported that their purchasing decisions are based on online product evaluations. Researchers study various research approaches to investigate the eWOM phenomena. They can classify eWOM communication in two levels: Market-level analysis and Individual-level analysis [9]. At the market-level analysis researchers focused on market-level parameters for example price [10]. At the individual level analysis, researchers supposed eWOM as a process of personal influence that senders can influence on receivers and change receiver's idea and purchasing decision [11].

There are two major problem in influence diffusion, first influence spread is unobservable in word of mouth network, so the tracing of influence is problem [12][13]. Second when the status of influence is successful, the observation of influence data is heavy [14]. Also one of the challenge in social advertisings increase enhancement and broad diffusion. In the previous studies, researchers considered on influential users who diffuse information, their followers have interest to this information finally they can maximize diffusion in social networks. In social advertising, identifying influential users is important issues because companies can reduce cost of advertising by them. But the influence of users may be different in diversity domain specificity. Users have strong influence in a topic and may don't have as influence in another topics. In this paper, we present a proposed method that called DMIU for discovering influential users based on domain specificity.

In the following of paper we would show, the users have same number of follower, they do not have necessarily same influence on other followers. We calculate influence base on the observable activity of users for instance (their content, their retweet). For this doing, we used Twitter application program interface, we compare our proposed method with other methods and showed improving performance rather than other models.

The rest of this paper is organized as follows. In the next section, some of the most important related works are reviewed. In section 3, the structure of proposed method for discovering influential users in social media .The evaluation of this method is discussed in section 4. Finally, last section presented conclusion.

II. RELATED WORK

By the emergence of social networks in recent years, finding in Influential users has absorbed a considerable amount of attention from researchers in this area. In this section, we review the literature separately on finding influential users on SNWs.

A large collection of weblogs are studied for transmission posts between bloggers. They peruse post's time-stamped and diffuse them between bloggers and showed the diffusion of posts followed independent cascade model [15]. Bloggers studied for epidemic of interest. They investigate that similar blogs use from the content of each other or not and pay attention that influence of blogger in another blogger can be with one post. They utilize a novel inference scheme that takes advantage of data describing historical, repeating patterns of "infection." [16]. Electronic-commerce network studied for propagation of recommendations and the cascade sizes. They analyze how user behavior varies within user communities defined by a recommendation network [17]. A limitation of these studies was the lack of data from structural of diffusion network. Recent studies pay attention to diffusion data and structural of diffusion network. Facebook social network studied to analyze diffusion chains in Facebook user's profile, they showed that after

controlling for distribution effects, there is no meaningful evidence that a start node's maximum diffusion Chain length can be predicted with the user's demographics or Facebook usage characteristics (including the user's Number of Facebook friends). [18]. Second life online game studied for information diffusion of "gestures" between friends and showed the transmission of information between two friends is more than two strange users. Also they showed that some users have important role more than another users called adopter [19]. Customer networks studied for discovering potential customers. They propose to model also the customer's network value: the expected profit from sales to other customers she may influence to buy, the customers those may influence, and so on recursively [20]. Choosing set of good customers in viral marketing studied for optimization problem. They divided customers to two categories that called active and passive. They proved that the optimization problem was NP-hard for both LT and IC Models. They presented the greedy algorithm. The greedy solution would be produced result which was optimal, but scalability was low in this method [21]. Water distribution network studied on sub modular property of influential function in Kemp's greedy algorithm and proposed CELF algorithm based lazy forward. In scalability, CELF was better than Kemp's greedy algorithm and memory usage of CELF was low than Kemp's greedy algorithm [22]. CELF++ proposed based on CELF and Two real world dataset from collaboration networks collected from arXiv (www.arXiv.com) showed that CELF++ algorithm was significant improvement in running time and number of node look up but the memory usage CELF++ was more than CELF[23]. Community-based approach proposed for reducing the computational cost of greedy algorithm of influence maximization. Call detail record (CDR) from China Mobile showed that the run time of CGA was faster than MixedGreedy and the influence spread of CGA was very close to MixedGreedy and NewGreedy[24]. SPINE model proposed for solving scalability problem with sparsification of network. SPINE had two phases, first selected a set of finite edges and second it greedily seek a solution of maximum. Yahoo! Meme and a prominent online news site showed that SPINE can reduce time running of influence maximization problem and number of active nodes in sparsification of network close to full of network [25]. Call records studied centrality measures in selecting influential customers that include of degree centrality, hubs centrality and page rank centrality [26]. Recent researches show that such action log can provide traces of influence among users in a social network Myspace network studied for the dynamics of the influence of users across different topics based on three measures: the number of followers, re-tweets, and mentions on Twitter network [24]. Twitter network Studied the influence of Twitter users by Number of followers, friends, and tweets Past influence of users [25]. Review Website studied for developing a model to estimate the influence capability of reviewers online. This model includes of Number of subjective terms, review frequency [26]. Facebook studied for the characteristics of social marketing messages. They analyzed the messages posted by restaurants with the most Facebook fans messages that contributed to different levels of popularity. Used the number of "likes" to measure the popularity of a message [27].

Previous studies have quantified influence in terms of network metrics for instance number of retweet or page rank, our proposed method (DMIU) measure influence in terms of the size Tree Regression. The focus of previous studies were based on the structural of diffusion and feature of content but Influential users have different influence in diversity domain specificity for instance in a special topic have strong influence and another topics have weak influence. So DMIU presented for identifying influential users based on domain specificity.

III. PROPOSED METHOD

Influential users have different influence in diversity domain specificity. For instance in a special topic have strong influence and another topics have weak influence. So DMIU method presented for identifying influential users based on domain specificity in this paper. The DMIU method will identify the influential users for delivering advertisements and it maximized the diffusion of information between the users of social network. The DMIU method takes advantage of content relevance and social relationships to reduce the negative impression of the advertisement and gain marketing effectiveness. It also suggests appropriate friends to users to share the information, which enhances the resonance and reduces the problem of social spam. in Figurer 1 showed the flowchart of DMIU method.

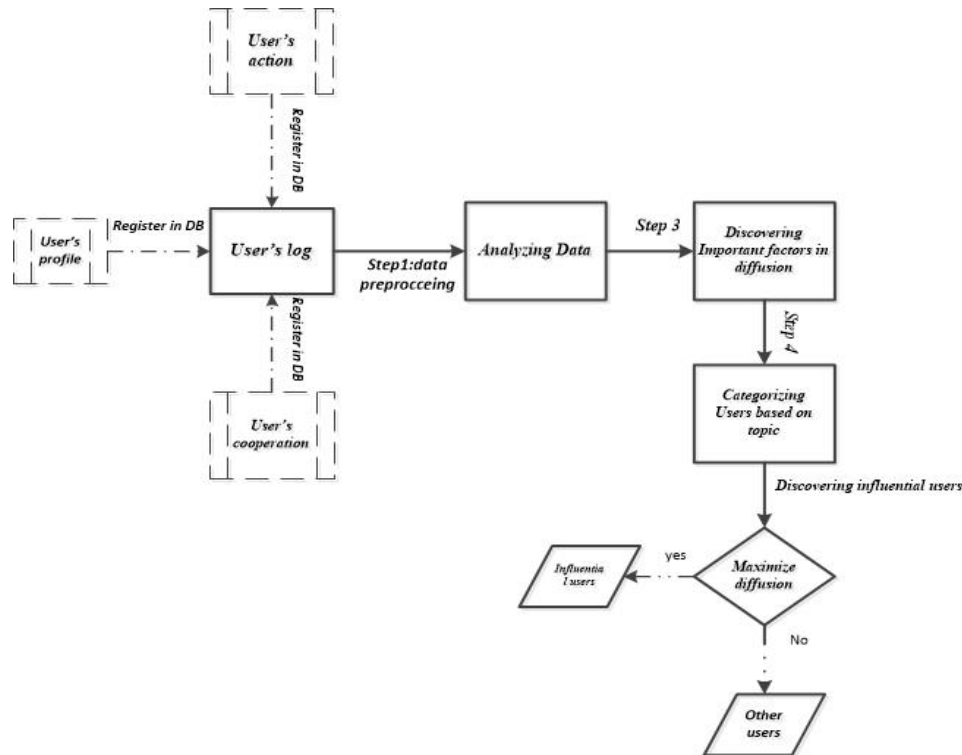


FIGURE. 1 THE FLOWCHART OF DMIU METHOD

First Step: Data Preprocessing

In social network, users do different actions such as joint to their interest group, find their friends, and follow their interest subjects. User's log includes of all these actions that divided to structural, semi structural and unstructured data.

Structural data resides in a fixed field within a record or file [28]. Structural data has the advantage of being easily entered, stored, queried and analyzed. In the real world, data are incomplete, noisy and Inconsistent. The task of data preprocessing includes of Data cleaning, data integration, transformation and reduction [28]. Unstructured data is all those things that can't be so readily classified and fit into a field same as images [29]. Semi structural data is between structural data unstructured data [30]. The preprocess of unstructured data showed in figure 2. Text cleaning Include of stemming, removal of punctuations, removal of expressions, split attached words, removal of URLs, escaping HTML characteristic and decoding data.

Second Step: Important Features in Improving Diffusion

Users interact with each other on social networks. These interactions include of like, mention, share, hashtag and another actions that can improve diffusion on social networks. The influence of these actions are different in diffusion, some actions have influence score more than another actions in diffusion. multiple-regression analysis is an appropriate method for developing a prediction model and analyzing the relationship between the actions and influence score, So multiple regression used for prediction score influence of actions for DMIU method in equation (1) [31]:

$$y = b_0 + b_1(x_1) + b_2(x_2) + \dots \quad (1)$$

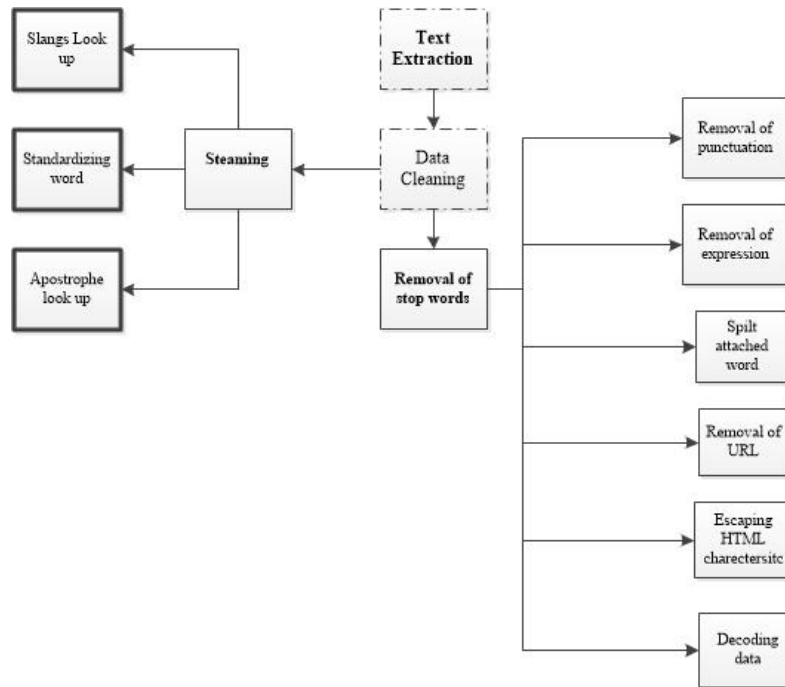


FIGURE. 2 SEMI-STRUCTURED AND UN STRUCTURED DATA PRE-PROCESSING

Y is response variable, x_p is predictor variables and b_p is clustering coefficient between response variable and every predictor variables in equation [31] (2):

$$\left(\frac{r_{y,x_p} - r_{yx_{p+1}} r_{x_p x_{p+1}}}{1 - (r_{x_p x_{p+1}})^2} \right) \left(\frac{SD_x}{SD_y} \right) \quad (2)$$

Third Step: Categorized User's Content Based On Domain Specificity

Users generate diversity contents and share them between friends .their follower showed reaction to these contents, contents include of different topics. In this section, user's contents categorized based on domain specificity. For this doing has two steps, first step, we measure important a keyword to a content by Frequency-inverse document frequency (TF-IDF) term that it is a numerical statistic that is intended to reflect how important a word is to a document in a collection. The term frequency (TF) for term m in a post p is calculated [32] as (3) [32], Where $\text{freq}_{m,p}$ is the raw frequency of term i appearing in post p and $\max_1(\text{freq}_{1,p})$ is the number of times the most frequent index term, l , appears in post j .

$$tf_{m,p} = \frac{\text{freq}_{m,p}}{\max_1(\text{freq}_{1,p})} \quad (3)$$

The inverse document frequency (IDF) for term m is formulated [32] as (4), Where N_p is the total number of posts and n_m is the number of posts in which term m appears.

$$idf_m = \frac{\log N_p}{\log N_m} \quad (4)$$

Then the relative importance of term m to post p can be obtained by calculating [32] as (5):

$$w_{m,p} = tf_{m,p} \times idf_m \quad (5)$$

Second step, we measure the similarity in the aspect of those keywords and domain specificity by Cramer clustering coefficient metric (6) and (7) [32]. K is number of rows, l is number of columns, F_o is real frequent, F_e is frequent:

$$v = \frac{\sqrt{\chi^2}}{\sqrt{n * (k - 1)(l - 1)}} \quad (6)$$

$$\chi^2 = \frac{(F_o - F_e)^2}{F_e} \quad (7)$$

Cramer's V equals 0 when there is no relationship between the two variables, and generally has a maximum value of 1.

Fourth Stage: Discovering Influential Users

The goal of this stage identifies influential users based on domain specificity. To identify influential users, we computed influence score of user's actions in second step and categorized user's content in third step, in fourth step we fit regression tree model based on greedy optimization process recursively partitions the feature space since regression tree model is much better calibrated than the linear regression model. For every domain specificity, regression tree compute feature space that include of conditions based on user's action and identify influential users, regression tree models much better calibrated than the linear regression model. in figure 3 shows the regression tree for one of the domain specificity. In this example three user's actions are important in influence score of diffusion than another user's actions that include of number of followers ($>\mu$) the time of diffusion ($>\pi$) and number of posts ($>\alpha$). these user's action are constraints in regression tree, where the left (right) child is followed if the condition is satisfied (violated). Leaf nodes give the predicted influence the corresponding partition. In Figure 3 For instance a user generate more than 60 posts and they diffuse information less than 5 minuet and they influence more than 270 followers, almost certainly they are influential users with probability 0.97.

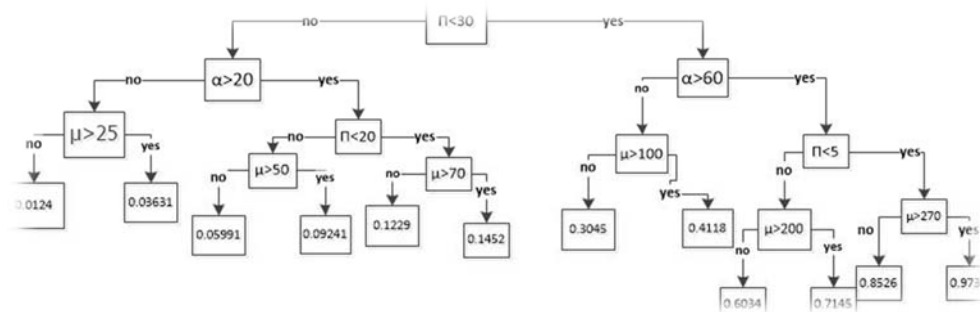


FIGURE. 3 THE EXAMPLE OF REGRESSION TREE FOR DISCOVERING INFLUENTIAL USER

IV. EVALUATION

In this section, we discuss the results of experimental study and some insights discovered from the observations and analysis. In order to evaluate the DMIU method on Twitter offer application program interface (API)². 13500 tweets posted from 420 users that selected randomly, for each public tweets the following information is average number of like, average number of Share, average number of comment and average number of mention. The statistical information of twitter API is depicted in table 1.

TABLE .1 SUMMARIES OF PROPERTIES OF TWITTER DATASET

Parameters name	Value of parameters
Number of users	420
Number of contents	13500
Average number of like	39150
Average number of Share	22950
Average number of comment	35100
Average number of mention	28650
Average number of friends	33615

Performance Analysis

The goal of our experiments is to appear that influence spread achieved by DMIU method improves influence spreads that can be achieved by approaches like [26][27]. We compared influence spread, number of nodes activated by DMIU method with method [26] and method [27].

Method [26]: this is a framework that combined with mining techniques, a modified point-wise mutual information PMI measure and recency, frequency, and monetary adaptive (RFM model) is proposed to evaluate the influential power of online reviewers. The features of this method include of Number of subjective terms, review recency and frequency, this method can use on-line word of mouth marketing.

Method [27]: this is a framework that used from text mining and statistics method to discovered influential popularity of social marketing message, the features of this method include of content and also considered the media type, whether “status,” “link,” “video,” or “photo,” of the message. The popularity of a message measured by number of like.

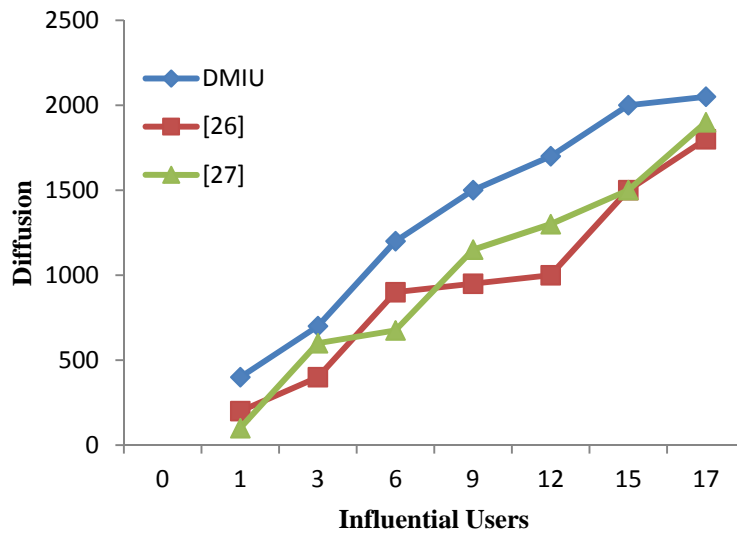
The user’s content of Twitter (API) categorized based on domain specificity in DMIU method that political topics, technology topics and sport topics. Influential users discovered in the every topic and determined influence of diffusion by them

Influential Users In Political Topics

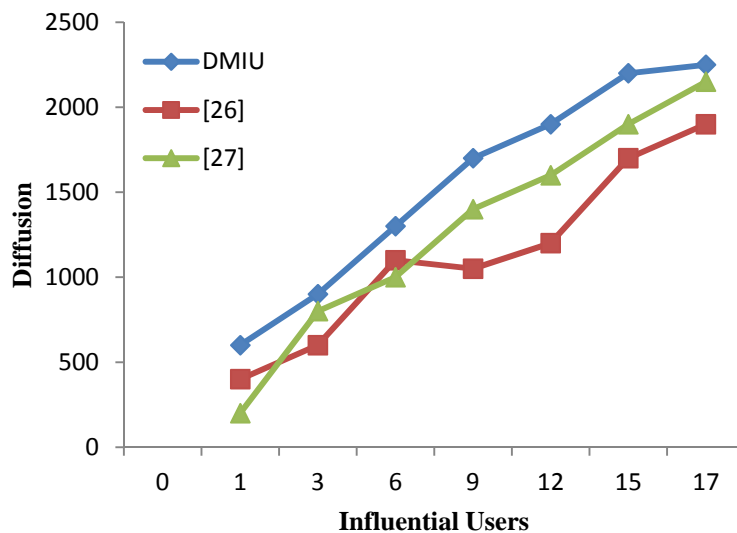
The number of Influential users is seventeen in political topics. In Figure 4.a (linear threshold model) the influence spread of DMIU method is more than method [26] and method [27]. In Figure 4.b (cascade model) when the number of influential users is smaller 6, DMIU method is closely to method [26] and method [27] but when the number of

² <http://snap.stanford.edu/data/index.html>

influential users are more than 6 the influence spread of DMIU method is more than method [26] and method [27].method [27] performs inconsistently compared with method [26] in linear threshold method and cascade method for political topics.



A. INFLUENCE SPREADS OF DIFFERENT METHOD ON POLITICAL TOPICS IN LINEAR THRESHOLD DIFFUSION

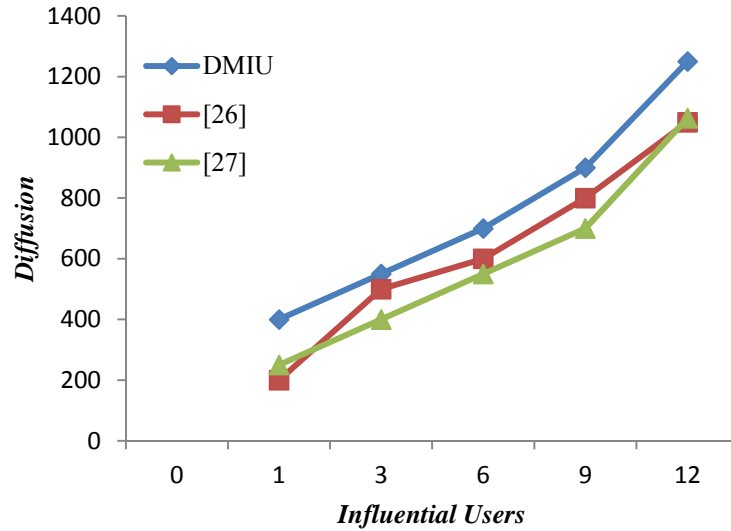


B. INFLUENCE SPREADS OF DIFFERENT METHOD ON POLITICAL TOPICS IN CASCADE DIFFUSION

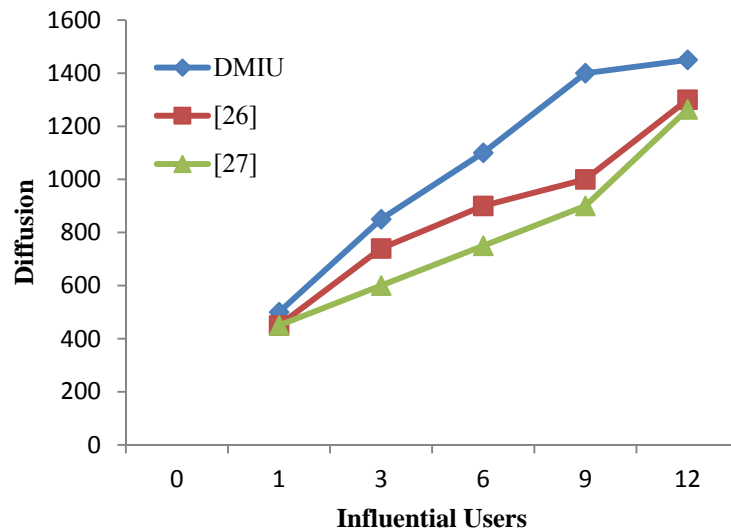
FIGURE 4. INFLUENCE SPREADS OF METHOD [26], METHOD [27] AND DMIU METHOD ON TWITTER DATASET

Influential Users In Technology Topics

The number of Influential users is twelve in technology topic. In Figure 5.a (linear threshold model), the influence spread of DMIU method for influential users smaller 3 is closely to method [27]. in Figure 5.b (cascade model) the influence of DMIU method is more than method [26] and method [27].method [27] increase influence of diffusion more than method [26] for technology topics in consider with political sports, in method [27] popularity of content can increase influence of diffusion rather than method [26].



a. INFLUENCE SPREADS OF DIFFERENT METHOD ON TECHNOLOGY TOPICS IN LINEAR THRESHOLD DIFFUSION

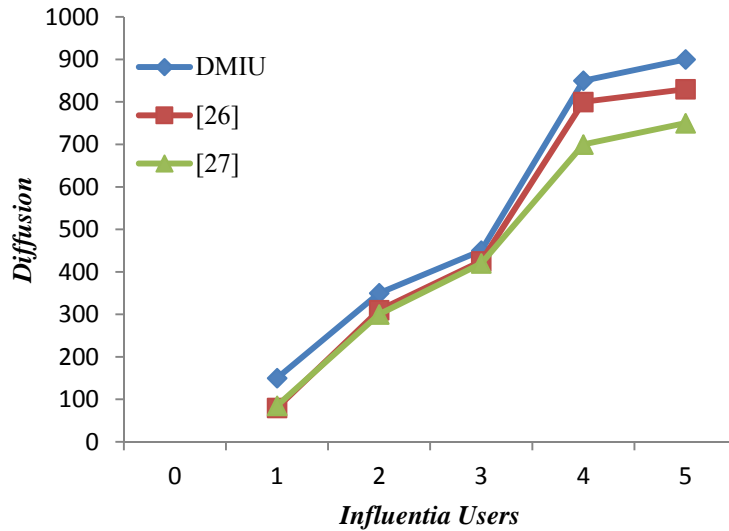


b. INFLUENCE SPREADS OF DIFFERENT METHOD ON TECHNOLOGY TOPICS IN CASCADE DIFFUSION

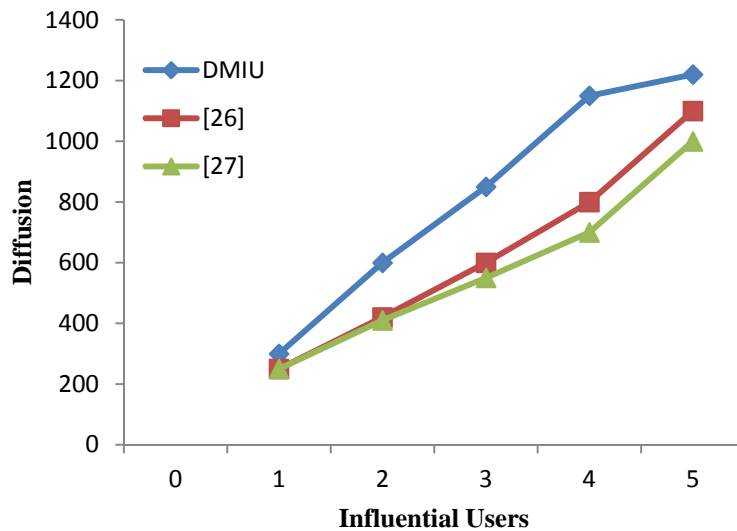
FIGURE 5. INFLUENCE SPREADS OF METHOD [26], METHOD [27] AND DMIU METHOD ON TWITTER DATASET

Influential Users In Sport Topics

The number of Influential users is five in sport topic. In Figure 6.a (linear threshold model), the influence spread of DMIU method is more than method [26] and method [27]. In Figure 6.b (cascade model) the influence of DMIU method is more than method [26] and method [27]. in (cascade model) for influential users smaller than 3 the influence spread of method [27] is closely to method [26] and the influence spread of DMIU method is more than method [26] and method [27].



A. INFLUENCE SPREADS OF DIFFERENT METHOD ON TECHNOLOGY TOPICS IN LINEAR THRESHOLD DIFFUSION



B. INFLUENCE SPREADS OF DIFFERENT METHOD ON TECHNOLOGY TOPICS IN CASCADE DIFFUSION

FIGURE.6. INFLUENCE SPREADS OF METHOD [26], METHOD [27] AND DMIU METHOD ON TWITTER DATASET

Accuracy And F-Measure Analysis

We use accuracy and F1-measure to evaluate the performance of the method. The accuracy is the ratio of the number of correct classifications to the total number of classifications [32] as (8):

$$\text{Accuracy} = \frac{\#TP + \#TN}{\#TP + \#FP + \#FN + \#TN} \quad (8)$$

The F1-measure is the harmonic mean between precision and recall. Assume the recall of class i is R_i and the precision of class i is P_i [36] as (9):

$$f_i = \frac{2R_i P_i}{(P_i + R_i)} \quad (9)$$

The experimental results are shown in Table 2. The results were obtained by 10-fold cross-validation. The original dataset was randomly partitioned into ten subsets. In each test, nine of the ten subsets were used as the training data to produce the method, and the remaining data was used to evaluate the method. The process was repeated ten times, and each subset was used as test data exactly once.

TABLE.2 ACCURACY AND F1-MEASURE OF THE DMIU METHOD, METHOD [24] AND METHOD [25]

	method [26]	method [27]	DMIU method
Accuracy	0.647059	0.65098	0.692157
f₁-measuer	0.784373	0.773722	0.802539
f₂-measuer	0.322371	0.365534	0.396531
f₃-measuer	0.315158	0.460148	0.587279
F1-measuer	0.511370	0.547989	0.621495

The results show that our DMIU method gives better and more robust prediction than method [26] and method [27]. The accuracy of DMIU method was 0.69 and method [26] and method [27] in order were 0.64 and 0.65.

V. CONCLUSION

Social influence has diversity application in real worlds for example electronic word of mouth marketing, social advertising and public opinion monitoring. In this paper, We argue that influential users in diversity domain specificity have not same influential on other users, they may have strong in influential a domain specificity and have weak influential in other domain specificity. This was motivated for presenting a method for identifying influential users based on domain specificity that called DMIU.

When users initiated to activity in these social networks, user's profile and user's actions (e.g. retweet) that influence on diffusion determined by "multiple regression" and user's contents categorized based on keywords by "TF-IDF" and finally influential users identified by "Tree Regression" based on domain specificity in this method.

Twitter (API) dataset used for evaluation of this method, three domain specify generated based on user's content that include of political topics, technology topics and sport topics. Influential users discovered based on every topic separately. DMIU method consider with method [26] and method [27].we showed the scale of diffusion in DMIU method is more than method [26] and method [27]. Also the accuracy of DMIU method and F1-measuer of DMIU method are better than another two methods.

Furthermore in future we wish to discover influential users in considering dynamic network. One key property of any social network is that it is changing all the time. Also we will use other parameters that may influence on diffusion for example cultural, racial, ethnic, and socioeconomic backgrounds.

VI. REFERENCE

- [1] Litvin, Stephen W., Ronald E. Goldsmith, and Bing Pan. "Electronic word-of-mouth in hospitality and tourism management." *Tourism management* 29, no. 3 (2008): 458-468.
- [2] Richins, Marsha L. "WORD OF MOUTH COMMUNICATION A NEGATIVE INFORMATION." *Advances in consumer research* 11, no. 1 (1984).
- [3] Chan, Yolanda YY, and Eric WT Ngai. "Conceptualising electronic word of mouth activity: An input-process-output perspective." *Marketing Intelligence & Planning* 29, no. 5 (2011): 488-516.
- [4] Cheung, Christy MK, and Dimple R. Thadani. "The State of Electronic Word-Of-Mouth Research: A Literature Analysis." In *PACIS*, p. 151. 2010.
- [5] Hung, Kineta H., and Stella Yiyan Li. "The influence of eWOM on virtual consumer communities: Social capital, consumer learning, and behavioral outcomes." *Journal of Advertising Research* 47, no. 4 (2007): 485-495.
- [6] Chang, Aihwa, Sara H. Hsieh, and Timmy H. Tseng. "Online brand community response to negative brand events: the role of group eWOM." *Internet Research* 23, no. 4 (2013): 486-506.
- [7] Lee, Mira, and Seounmi Youn. "Electronic word of mouth (eWOM) How eWOM platforms influence consumer product judgement." *International Journal of Advertising* 28, no. 3 (2009): 473-499.
- [8] Chevalier, Judith A., and Dina Mayzlin. "The effect of word of mouth on sales: Online book reviews." *Journal of marketing research* 43, no. 3 (2006): 345-354..
- [9] Lee, Mira, and Seounmi Youn. "Electronic word of mouth (eWOM) How eWOM platforms influence consumer product judgement." *International Journal of Advertising* 28, no. 3 (2009): 473-499.
- [10] Dellarocas, Chrysanthos, Xiaoquan Michael Zhang, and Neveen F. Awad. "Exploring the value of online product reviews in forecasting sales: The case of motion pictures." *Journal of Interactive marketing* 21, no. 4 (2007): 23-45.
- [11] Park, Do-Hyung, and Jumin Lee. "eWOM overload and its effect on consumer behavioral intention depending on consumer involvement." *Electronic Commerce Research and Applications* 7, no. 4 (2009): 386-398.
- [12] Liben-Nowell, David, and Jon Kleinberg. "Tracing information flow on a global scale using Internet chain-letter data." *Proceedings of the National Academy of Sciences* 105, no. 12 (2008): 4633-4638.
- [13] Snow, Rion, Brendan O'Connor, Daniel Jurafsky, and Andrew Y. Ng. "Cheap and fast---but is it good?: evaluating non-expert annotations for natural language tasks." In *Proceedings of the conference on empirical methods in natural language processing*, pp. 254-263. Association for Computational Linguistics, 2008.
- [14] Dawes, Robyn M. *Everyday irrationality: How pseudo-scientists, lunatics, and the rest of us systematically fail to think rationally*. Westview Press, 2001.
- [15] Gruhl, Daniel, Ramanathan Guha, David Liben-Nowell, and Andrew Tomkins. "Information diffusion through blogspace." In *Proceedings of the 13th international conference on World Wide Web*, pp. 491-501. ACM, 2004.

- [16] Adar, Eytan, and Lada A. Adamic. "Tracking information epidemics in blogspace." In *Proceedings of the 2005 IEEE/WIC/ACM international conference on web intelligence*, pp. 207-214. IEEE Computer Society, 2005.
- [17] Sisovsky, Karel. "How does marketing become viral?: An investigation of the dynamics in digital campaigns." (2015).
- [18] Sun, Eric, Itamar Rosenn, Cameron Marlow, and Thomas M. Lento. "Gesundheit! Modeling Contagion through Facebook News Feed." *InICWSM*. 2009.
- [19] Bakshy, Eytan, Brian Karrer, and Lada A. Adamic. "Social influence and the diffusion of user-created content." In *Proceedings of the 10th ACM conference on Electronic commerce*, pp. 325-334. ACM, 2009.
- [20] Domingos, Pedro, and Matt Richardson. "Mining the network value of customers." In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 57-66. ACM, 2001.
- [21] Kempe, David, Jon Kleinberg, and Éva Tardos. "Maximizing the spread of influence through a social network." In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 137-146. ACM, 2003.
- [22] Leskovec, Jure, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie Glance. "Cost-effective outbreak detection in networks." In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 420-429. ACM, 2007.
- [23] Goyal, Amit, Wei Lu, and Laks VS Lakshmanan. "Celf++: optimizing the greedy algorithm for influence maximization in social networks." In *Proceedings of the 20th international conference companion on World wide web*, pp. 47-48. ACM, 2011.
- [24] Wang, Yu, Gao Cong, Guojie Song, and Kunqing Xie. "Community-based greedy algorithm for mining top-k influential nodes in mobile social networks." In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1039-1048. ACM, 2010.
- [25] Xu, Wen, Weili Wu, Lidan Fan, Zaixin Lu, and Ding-Zhu Du. "Influence Diffusion in Social Networks." In *Optimization in Science and Engineering*, pp. 567-581. Springer New York, 2014.
- [26] Li, Yung-Ming, Chia-Hao Lin, and Cheng-Yang Lai. "Identifying influential reviewers for word-of-mouth marketing." *Electronic Commerce Research and Applications* 9, no. 4 (2010): 294-304.
- [27] Yu, Bei, Miao Chen, and Linchi Kwok. "Toward predicting popularity of social marketing messages." In *Social Computing, Behavioral-Cultural Modeling and Prediction*, pp. 317-324. Springer Berlin Heidelberg, 2011.
- [28] The Penn database group has structured and XML data project
- [29] Bevington, P. R., and D. K. Robinson. "McGraw-Hill; New York: 2003." *Data Reduction and Error Analysis for the Physical Sciences*: 212-216.
- [30] Structure, Models. "Meaning: Is" unstructured" data merely unmodeled." *Intelligent Enterprise*, March 1 (2005).
- [31] Finkelstein, Michael O. "The judicial reception of multiple regression studies in race and sex discrimination cases." *Columbia Law Review* 80, no. 4 (1980): 737-754..
- [32] Kjaerulff, Thora M., Francisco Rivera, Antonia Jiménez-Iglesias, and Carmen Moreno. "Perceived quality of social relations and frequent drunkenness: a cross-sectional study of Spanish adolescents." *Alcohol and Alcoholism* 49, no. 4 (2014): 466-471.
- [33] Salton, Gerard. "Developments in automatic text retrieval." *Science* 253, no. 5023 (1991): 974-980.

StudentPIMS: A Smartphone-Based Personal Information Management System for Students

Irfan Ullah^{a,b}, Shah Khusro^b, Habib Un Nabi^a, Rafi Ullah^a

^aDepartment of Computer Science, Shaheed Benazir Bhutto University, Sheringal, 18050, Pakistan

^bDepartment of Computer Science, University of Peshawar, Peshawar, 25120, Pakistan

Abstract—Curricular and co-curricular activities are among the major responsibilities that require proper attention from the students in order to achieve different goals and objectives regarding their bright future. Because of the mismanagement of keeping personal information about these activities, most of students are unable to remember these tasks while they are busy in their studies and therefore, fail to perform these activities at the right time. To handle this issue, they adopt several means including SMS drafts, reminders, sticky notes, notebooks, dairies, and laptops etc., which are limited and unable to fully support students because of several problems including their storage, search, and retrieval. With the availability and wide-spread adaptation of Android and Smartphones, researchers and developers started thinking of new and innovative ways of managing personal information of people especially students. Today, several apps are available on Google Play for managing personal information of students. However, the existing solutions have limitations including bulky user interfaces especially when the stored information exceeds a certain limit, usability, privacy, and requiring access to Internet for accessing certain services, which becomes a barrier to students especially to those living in rural areas of developing countries where access to Internet is among the major issues. Keeping in view these limitations, we have designed and developed StudentPIMS - a simple and usable Android app that allows students to easily manage personal information about these activities without suffering from cognitive overload caused by existing solutions. We have compared our solution with the existing solutions using some evaluation metrics as well as conducted a survey research among users of the app. Results show that StudentPIMS outperforms the available solutions especially in terms of usability, privacy, and low resource consumption.

I. INTRODUCTION

Personal Information Management (PIM) is the combination of theory and practice of recording, organizing, maintaining, searching, retrieval and use of items related to personal information such as short text messages (SMS), emails, personal documents, web pages, and reminders stored in the form of sticky notes or short text in mobile phones. These items help users in performing different tasks related to their day-to-day life [1]. It is concerned with methods, tools, and techniques which people adopt to keep and manage personal information available in varying settings with several purposes and with varying nature of information. For example, a user may manage their printed files in shelves with a proper alphabetical order and their digital documents in folders with proper nomenclature so that they can remember the name of the folder in which they stored a digital document [2]. We often divide the space on a hard drive in different logical partitions and then add some folders with proper names, which allow keeping our songs, videos, pictures and other items organized, and help us in easily finding these items when required.

For managing personal information collections, users adopt different methods ranging from writing on printed papers such as notebook to digital devices including PCs, mobile phones and Smartphone [1]. Everyone knows the well-known reminder feature in ordinary mobile phones. When it comes to student life, they have to remember information regarding their curricular and co-curricular activities, which are quite important because education is the most important objective in teen-age and university life. These tasks can be performed well if students are reminded in time about important activities. Students are often confronted with tight schedule and a batch of activities to deal

with. For this purpose, students may use sticky notes in printed and digital form, reminders in their mobile phones, desktop PCs and laptops, or write in notebooks. However, all these have some limitations. For example, desktop PCs and laptops cannot remind you especially when they are off, reminders lack alarming systems, and sticky notes may get deleted or lost. Although some Smartphone-based apps have been designed but these are either complex, need Internet connection or lack reminding students at proper occasions. In order to cope with these issues, we need to develop a Smartphone-based student personal information management system that helps students in managing their personal information easily and effectively. Today, with the wider adoptability of Android and Smartphones, it has become possible to develop low-cost Android apps for managing personal information. We are bringing your attention to the issue of PIM for students and therefore, propose StudentPIMS, a PIM system for managing information about students' curricular and co-curricular activities that keep students reminded with its alarming system so that these tasks are not missed. Therefore, the main objectives of this paper include:

- To investigate and briefly review the state-of-the-art Android-based Personal Information Management Systems (PIMS) that have been developed for managing and keeping record of different academic activities of students and their teachers
- To develop Android-based PIMS for students (StudentPIMS) that can manage and keep track of their curricular activities like assignments, presentations, tests, and quizzes and other co-curricular activities along with due time and date so that students can be alarmed before the actual event takes place.

Along with students, StudentPIMS can also be adapted to augment the memory of their teachers in keeping them remembered about different activities including checking assignments, taking planned and surprise quizzes and tests and preparing research articles and lectures for conferences, journals, and students respectively. This way, the app will make the academic life of both students and their teachers much easier. Rest of the paper is organized as Section-II covers literature review and related work, Section-III presents the architecture, user interface, and implementation details of StudentPIMS, Section-IV presents results & discussions on the evaluation of StudentPIMS, and finally Section-V concludes our discussion and places some future directions before researchers and developers. References are presented at the end.

II. LITERATURE REVIEW

The term personal information management is concerned with handling personal information, but first we need to understand what personal information is. It is the information that is under the control of the person and is kept by a person directly in their mind or indirectly e.g., in a diary, notebooks or through software applications. It is also the information that is about a person but not under their control and kept by others including their doctors, hospitals, and government agencies etc. It can also be the information that the person experiences but cannot control it. This information includes the examples of reading books and visiting websites etc. [3]. Now Personal Information Management (PIM) can be defined as a system that an individual creates or develops for their personal use in their work environment. Such a system may include methods and rules for acquiring, organizing, storing, maintaining and retrieving information along with procedures that produce several outputs [3]. In simple words, the activities of PIM map the user needs with the required information [3].

A. General Approaches towards PIM

Among the major activities of managing personal information, the most important problem for people is how to successfully find personal information because it is typically a complex process consisting of multiple steps, even when they know very well the information target. Many factors affect this task including the type of information, the information seeker, and the task of finding. The finding/re-finding of personal information on one hand is very similar to finding new information, but can be very different on the other hand as it requires the additional knowledge of where the seeker has kept the required information. Other factors including how information was originally encountered, organized into a specific structure, and the change in information environment with the passage of time [3].

Besides storing personal information in the notebooks, dairies and in mobile phones in the form of reminders and SMS drafts, users (students) also use other digital solutions such as keeping personal information in emails, cloud storage, and in the form of digital documents organized into folders and subfolders. While the paper-based environments have their own issues, our focus here is on storing, organizing, and finding/re-finding personal information in digital environments such as email, the Web, PCs, and mobile phones etc. Among the tasks of managing personal information, the finding/re-finding of information is the most required task, which according to [3] is handled by several solutions like keyword searching, Google Desktop, Apple Spotlight, and Microsoft Desktop Search etc. [3]. However, mobile phone users may use reminders or save a task in the form of SMS draft. Similarly, a number of apps are also available on Google Play, which can be installed on Android-based Smartphones. In this regard, our focus in this paper is on PIM solutions that have been specifically designed for managing personal information of students related to their curricular and co-curricular activities.

B. Students' Practices towards PIM

People adopt varying behavior while managing their personal information. They may be good at keeping their digital collections like documents and emails well-organized but show disorganization in keeping their printed documents including bank statements, research papers and other formal documents [1]. Same could be the case of students while dealing with their personal information. For example, according to Mizrahi [4] students use their laptops and desktop PCs in keeping their course and study materials, check emails, and go on Social Web applications like Facebook, however, students still feel them as a hurdle in portability. Students also go to discussion forums on the Web while seeking solutions to their problems regarding courses and topics as well as regularly visit their course web pages in order to check new announcements, discussions, home works, assignments, lecture notes, presentation slides and recorded lectures. In order to communicate with class fellows, teachers, and supervisors, they use email as the primary source of communication while for communicating with friends they use instant messaging, texting (SMS), online chatting, making phone calls, and using Social Web applications like Facebook, Twitter etc. [4].

Hashemzadeh and Salehnejad [5] conducted a more comprehensive study reporting PIM behavior of students. They reported that computers and laptops are most widely adopted devices for managing personal information of students especially their course materials. External storage is ranked second, cell phones are ranked third, and the Web is ranked fourth in this list [5]. However, compared to cell phones, Smartphones are now widely used among

students and professors, and with proper care of data privacy, we believe that students will be using majority of apps available on Google Play website. Therefore, designing apps that securely and easily manage the personal information of students especially regarding curricular and co-curricular activities is of immense importance, which is the major goal of this research work.

C. Managing Data & Privacy on Mobile Phones

Mobile phone users are much curious about keeping their data safe and private, and therefore, may be reluctant to install an app or may reinstall an app when they come to know that their data privacy may be compromised. This is confirmed from a study conducted by Boyles, Smith, and Madden in 2012 [6] that reports that about 41% of mobile phone users keep a backup of their data including contacts, images and other necessary files in order to ensure if their phone is lost or its security is compromised. They further added that about 32% of the subjects regularly keep cleared their browsing and search history on their phones. About 19% of these phone owners keep off the location tracking feature on their mobile phones, as they don't want even to share this information. About 24% of the subjects reported that their privacy became a question when their phone was hacked by someone. Similarly, about 54% of the cell phone owners refused to install apps that want personal information up to a certain threshold [6].

This study conducted by Boyles et al [6] becomes a serious concern for students who want to install PIM type solutions on their phones. Therefore, security and privacy need to be ensured. In this regard, we propose that a PIM solution should only access phone resources (e.g., alarm, clock etc.) at lower hardware and software level through Android OS and should not access the personal content of the users including contacts, photos, and other files. Similarly, the app should not be accessible through Internet and the Web so that no hackers can hack the personal information stored in the app. The proposed StudentPIMS takes care of these aspects of data privacy.

D. Related Work

Although a number of PIM solutions are available today, still we are unable to properly manage a significant portion of information that subsequently becomes out of reach when we need them. This unmanaged information is often available in the form of sticky notes on paper pieces attached to sheets of book or notebook, to the corners of our room tables and walls, remains in our pockets for quite some time until lost or wasted, remains in email account folders, and archived in the form of digital documents in our laptops and Smartphones. These scattered pieces of information can be termed as information scraps, which never make it to the PIM applications [7]. Now-a-days, the wide-spread use of Smartphones and Android-based solutions, it has become possible to present a one-size-fits-all solution by keeping every information scrap at a single location, i.e., using Smartphone apps. Whenever, a student needs taking notes about a particular curricular or co-curricular activity, they may think of StudentPIMS so that they can retrieve it back whenever and wherever required.

According to Lorna et al [8], managing personal information is a major issue for general public as well as for faculty members, where information overload is among the major challenge. They surveyed that faculty members use email, desktop PCs, Web-based information systems, and learning management systems to keep organized their digital content [8]. Therefore, ways must be devised that resolve these issues, where designing a PIM solution may definitely reduce the cognitive and information overload in managing personal information up to a major extent. In this regard, a number of PIM apps have already been developed which are discussed in the coming paragraphs.

Power School [9] is Android app that provides up-to-date information to students and their parents about different aspects of student's progress including attendance, assignments, and test scores etc., using push notifications from the Power School web server. It can be used around the globe by students of any school that is subscribed to Power School Web information system; otherwise students will not be able to use it. Salient features include access to home works, assignments, grades, attendance, teacher comments, and one account for all children, daily bulletin board, and automatic and push notifications through email etc. [9]. It is limited as it suffers from login issues and requires new email address in case re-install is necessary after a Smartphone crash. Information about teachers cannot be updated. Similarly, reviews posted on its Google Play page confirm that the users are not satisfied with its working.

My Class Schedule [10] maintains a timetable of all the upcoming events including classes, examination, and unfinished work in a given day or week using its Web-based services. Other helpful features include grade overview of the student and notifying students of upcoming events through notifications. However, reviews posted on its Google Play page reveal that it suffers from a number of limitations including lacking some functionalities, problems in properly notifying students about due and upcoming events, availability of some features such as cloud storage and device synchronization on subscription of 1.90\$ [10]. Moreover, the app is not easy to use in managing the personal information of students especially about other co-curricular activities e.g., searching for a particular event or entry becomes difficult as users have to manually look up for a given event. Similarly, users have to distinguish between activities e.g., sports and classes, which are displayed next to each other. This involves a great deal of cognitive overload.

Students: Timetable [11] is another timetable-like Android app, designed to keep information about students' curricular activities through a timetable that is integrated to an organizer and a diary so that students can keep information and reminders about upcoming quizzes, tests, classes etc. [11]. However, the app has a number of limitations including no support for adding co-curricular activities, keeping limited information about teachers, and no support for searching and updating subject-related information. According to reviews posted on its Google Play page, users are complaining about too much ads popping up.

Student Agenda [12] is a lightweight and simple personal information manager for students that allows students to keep important information about student activities including curricular, co-curricular and daily life activities. Students can easily keep information about home works, tests, and appointments and get reminded whenever required. Timetable of events and tasks further augments user memory about different scheduled tasks [12]. However, the timetable becomes congested after adding a number of activities, where events are not categorized into curricular and co-curricular activities so their management becomes a problem. Therefore, the usability of the app also becomes a big question. Similarly, it accesses Smartphone camera and other personal information including photos etc., which can of great concern to students who are reluctant to installing apps as they feel threat to their privacy.

Student Buddy [13] helps in organizing student academic life and helps keeping them updated about upcoming events especially classes and lectures through notifications that can be set a specific time before the lecture. It also tells the users that how much classes they should attend to complete minimum attendance for a given subject [13].

However, it suffers from a number of limitations e.g., adding detailed information about activities, no categorization of events into curricular and co-curricular activities and searching for a given activity is difficult. In addition, it notifies the student one day before the assignment is to be submitted, where the student may have not completed the assignment. It would be great if students could dynamically set the alarm so that they can get reminded at the right time for doing the assignment along with the current functionality. The app needs further attention for its greater usability.

My Study Life [14] is a powerful PIM system designed for students that synchronizes student data in seamless manner between several devices so that users can easily add task on the fly which is made instantly available on the web app. Students can easily track their tasks including home works, assignments, presentations, and reminders, which are accessible anywhere and at any time. They can keep information about exams, classes and get notified about unfinished work [14]. However, it is not the case that a student will have multiple devices and will switch between them, and therefore, this synchronization may be problematic as it is time-consuming job and may require Internet to access the details about tasks from the cloud. This also introduces the issues of user privacy if the data is accidentally or intentionally compromised. Similarly, it lacks home screen widget so the users have to use the back button of the phone, which slows down the use of the app. It lack the calendar widget to connect an event with the calendar, no tracking of attendance and information about whether the student has taken the class or not etc.

Although several Android-based PIM solutions are available for students where the list is not limited to the one presented in this Section, however; these solutions suffer from a number of noteworthy limitations. For example, some of these suffer from login issues, poor usability, paid services, missing or incomplete functionalities such as inability to properly notify students, lack for distinguishing among curricular and co-curricular activities, limited browsing and searching options, limitations in keeping detailed information about teachers or their subjects, and the cognitive overload in getting to the desired activity especially when events are distributed on timetable. Moreover, as discussed in Section II that mobile phone users are much curious about keeping their data safe and private, and therefore, may be reluctant to install an app when they come to know that their data privacy may be compromised [6], and therefore, proper measures should be taken for keeping student personal information safe from being compromised. Therefore, we propose StudentPIMS (Section-III) that is a simple and easy to use PIM solution that keeps information about students' curricular and co-curricular activities in an easily manageable way along with taking into account the privacy of students by not accessing any of their personal data except the one that is stored in its database. We propose that StudentPIMS should be wireless and Internet-free solution so that its useful features can be taken advantage of to its fullest.

III. THE PROPOSED SOLUTION – STUDENTPIMS

In managing personal information of students, StudentPIMS should be able to keep track of both curricular and co-curricular activities. Students should be able to add, update, delete, view, browse, and search activities. The functional requirements, therefore, can be broadly divided into two categories namely: managing activities and viewing activities. Managing activities can be further divided into adding curricular activities, adding co-curricular activities, and viewing the activity log for already registered activities. Students can add curricular activity by selecting add curricular activity, where they select activity type such as test, assignment, presentation, and quiz etc.,

title, description and hints about the activity, teacher name, subject name, and set date and time so that they can be reminded at the right date and time. However, students are free if they don't want to set the alarm for an activity in which case they are reminded at least 12 hours before the due date and time. Note that subjects and teachers can be added at runtime as well so that students don't have to go back and forth for adding teachers and subjects. Students can also add co-curricular activities, where information like activity type such as party, game, tour etc., with all the necessary information such as title, description, venue and date & time. Similarly, students can view and update curricular and co-curricular activities through the activity log. Besides these functionalities, we are concerned with usability, privacy, and performance of StudentPIMS so that students can take advantage of the app to their fullest.

A. The Architecture of StudentPIMS

Figure 1 shows a three layered architecture of StudentPIMS consisting of user interface layer, application logic layer, and data layer. The user interface allows users to manage information about their curricular and co-curricular activities as well as view, browse, search and delete or update activities. The application logic layer is responsible for providing the necessary application logic that facilitates all the user supported activities on the user interface along with communication to the SQLite database and Android OS for necessary hardware and software support including setting alarm, using clock, and keyboard as well as using the hardware resources including screen and speaker etc. The main activity handler is responsible for communicating with user on the user interface and dealing with data and other resources on data layer. The data layer provides Android OS resources, hardware resources and access to SQLite database.

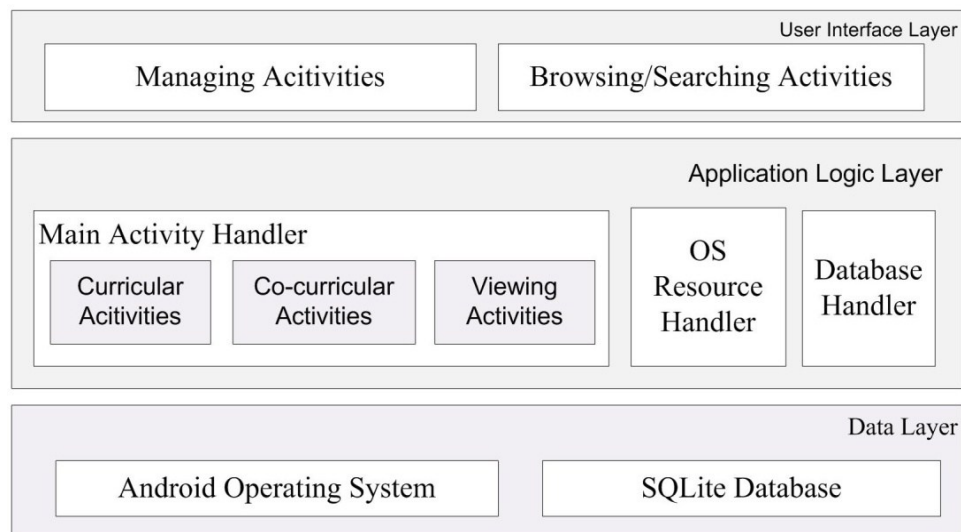


Figure 1. Architecture of StudentPIMS

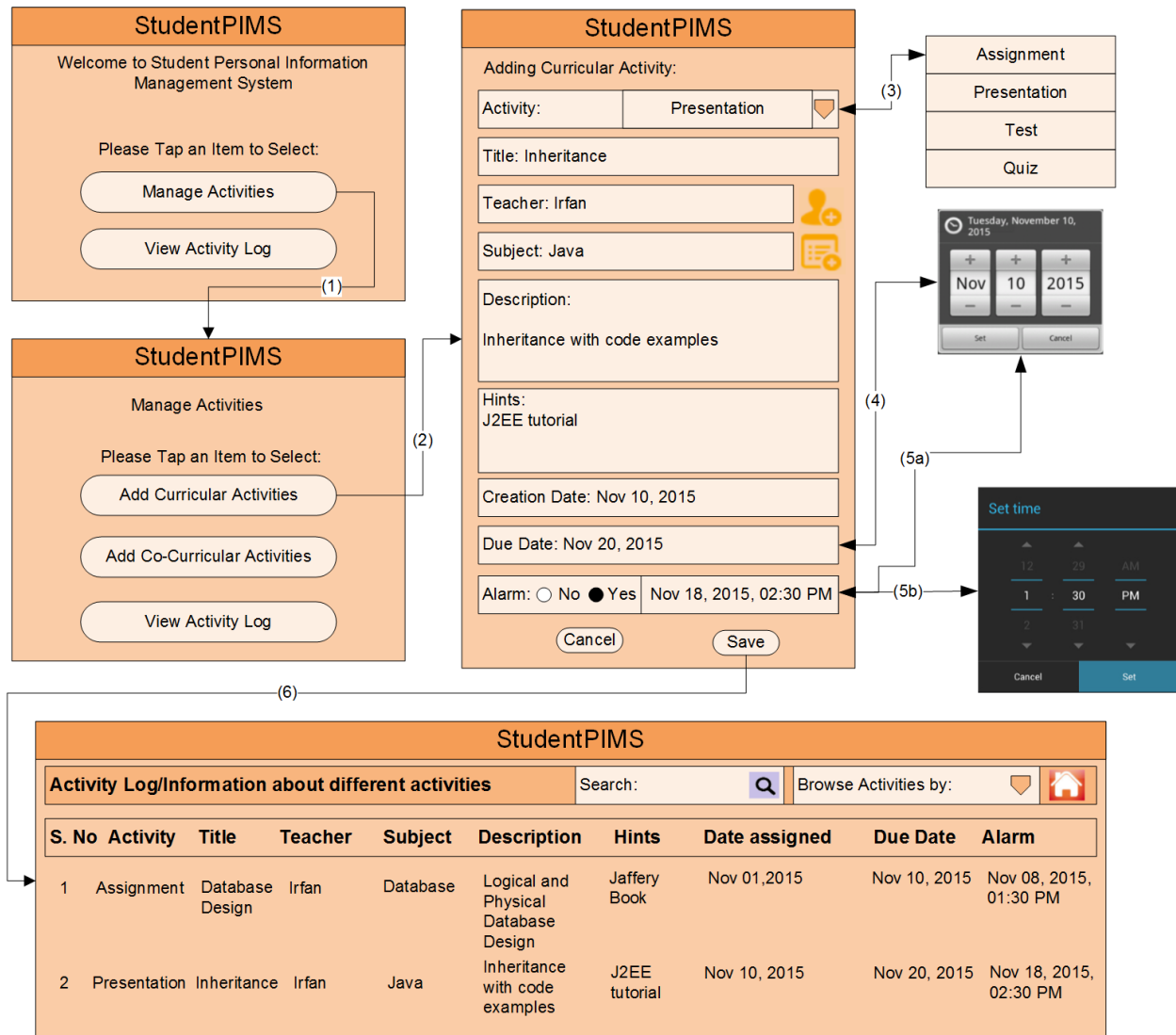


Figure 2. The proposed user interface for adding curricular activities

B. The User Interface

Figure 2 shows the proposed user interface of StudentPIMS for adding a curricular activity. After the welcome screen, the user selects manage activities leading the user to the next screen with three options: add curricular activity, add co-curricular activity, and view the activity log. Taping the add activity, a third screen appears where data about a curricular activity can be added. The user is required to select an activity from the drop-down list of activities including assignment, presentation, test, and quiz etc., title of the activity, teacher, subject, description, hints given by the teacher, due date, and setting alarm date and time. From this screen, the user can also add/remove teachers and subjects on the go by tapping the add teacher and add subject button. By taping the save button, the activity log is displayed showing the newly added item. In the similar way, co-curricular activities are added. Similarly, tapping view activity log allows users to view curricular or co-curricular activities. Note that on the activity log one can also add, delete, and update activity by pressing a particular record for long time. Users can also browse the records by teacher, subject, or activity type and also perform searching by entering keywords.

C. Implementation

Several hardware and software tools were used in the development of StudentPIMS. The hardware tools include Dell Inspiron Core i3 Laptop, model 5010 with 4GB RAM, 15 inches screen size, and 2.24GHz processing speed for installing development tools and documentation. Among the software tools, Android ADT version 22.3.0, Android SDK version 22.3.0, and JDK 7.0 update 79, Eclipse IDE Kepler version 1.0, XML 1.0, Microsoft Windows 7.0 professional, Microsoft Word 2010, and Microsoft Visio 7.0 are worth mentioning. Several Android OS libraries were used in order to access different resources of Smartphone for the smooth running of the application. Figure 3 depicts some of the snapshots of the proof-of-concept version of StudentPIMS on Huawei SCL-U31 Smartphone.



Figure 3. Our StudentPIMS app: Some of the snapshots of the proof-of-concept version of StudentPIMS, where (a) the loading screen, (b) welcome screen, (c) Adding co-curricular activity, (d) Adding curricular activity, (e) Activity log for curricular activities.

IV. RESULTS & DISCUSSIONS

Table I shows the evaluation of StudentPIMS against the available solutions using a few simple evaluation criteria namely usability, privacy, dependency on server & Internet, performance and the display of ads when connected to the Internet. Usability measures StudentPIMS against the available apps in terms of how much the app is simple and easy to use in accomplishing different tasks. Some of the apps while their installation on Google Play ask the user

to allow access to personal data including pictures, contacts, and other files, which is a great threat to privacy of the students. The metric privacy considers any app to be respecting of privacy if it is not accessing the personal files and is only limited to the information that added or deleted from its database. Dependency on server & Internet means that some apps require connectivity to their server that provides such services through Internet connections, which open way for hackers and malicious insiders to attack the personal files of user. Similarly, such dependency limits the apps in developing countries where connectivity to Internet and server is a major issue. The degree of details means how much information is stored in the database of the app. Here we are concerned with the apps keeping details of the curricular and co-curricular activities of the students which must be detailed so that the user takes full advantage of the reminders. Performance means the time taken by users in registering in activity, which will have two possible values: fast and slow. Some apps display ads when connected to Internet, which according to reviews from users on Google Play are teasing the users especially when ads display is frequent. In the light of this discussion, we conclude that StudentPIMS outperforms the available apps in terms of usability, privacy, and performance by allowing students to quickly and easily manage information about curricular and co-curricular activities. Similarly, it is free, free from ads display, and requires no specific resources such as Internet or web servers for its operation, and therefore, best in the developing countries where access to Internet and the Web is a major concern.

TABLE I
COMPARISON OF STUDENTPIMS AGAINST THE AVAILABLE PIM APPS

Evaluation Criteria \ Android Apps	Usability	Privacy	Dependency on Server & Internet	Subscription Mode	Performance	Ads Display
StudentPIMS	✓	✓	✗	Free	High	✗
Power School Mobile	✗	✗	✓	Subscription	Low	✓ (ad free version on subscription charges)
My Class Schedule: Timetable	✗	✗	✗	Free (fee for some features)	High	✗
Students-Timetable	✓	✗	✗	Free	High	✓
Student Buddy	✗	✓	✗	Free	Low	✗
My Study Life	✗	✗	✓	Free	High	✗
Student Agenda	✗	✗	✗	Free	Low	✓ (ad free version on subscription charges)

In order to further evaluate StudentPIMS, we installed the app on the tablets and Smartphones of 80 students and allowed them to use the app for about three months. The sample we selected consisting of 80 students including male and female students from different departments such as Computer Science, Forestry, Pharmacy, Sociology, Law and English. After using for about three months a questionnaire was distributed among them in printed format and students were requested to fill in the questionnaire so that we can further improve the design of the app. Among 80 students, 70 students responded by filling in the questionnaire. Among these 70 students, 61 were male and 9 were female students. We asked them 7 simple questions, which they answered on the questionnaire. These questions are given in Table II.

TABLE II
LIST OF QUESTIONS ASKED IN THE SURVEY OF STUDENTPIMS

S. No.	Questions
1.	How do you manage your personal information?
2.	You are satisfied with existing PIM solutions
3.	You are satisfied with StudentPIMS
4.	StudentPIMS is simple and easy to use
5.	You are satisfied with StudentPIMS privacy
6.	You recommend StudentPIMS to students
7.	StudentPIMS should be on Google Play

Figure 4 shows the statistics recorded in response to the first question. Note that cell phone reminders, SMS drafts, laptops, and notebook are mostly used in recording and managing students' personal information. However, if we look at the nature of these options, reminders, SMS drafts and students' personal notebooks are very limited solutions where managing and organizing personal information is very difficult. Although the use of laptops is a good choice, where a number of PIM solutions can be installed, however, their portability and power consumption can be problematic in several occasions. The use of Smartphone apps as PIM solutions is only about 8%, which is very low as compared to other solutions. This is because most of the students are either unaware of the existing PIM solutions or they are not satisfied with these solutions.

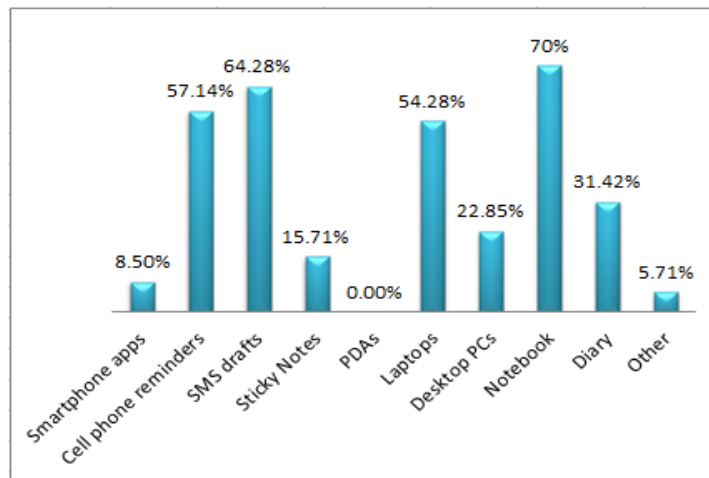


Figure 4. Distribution of students according to the type of devices used for PIM

Table III shows the statistics about students who are satisfied with existing PIM solutions. Among these, about 55% students (40% disagree, while 11.4% strongly disagree) are not in favor of existing solutions and only 19% students support the use of existing solutions. This confirms that some alternatives like StudentPIMS are needed to be developed. Similarly, the statistics shown in Table IV shows the satisfaction of students who have used StudentPIMS for three months. Here about 97% students (67 % strongly agree and 30% agree) are supporting the use of StudentPIMS, while remaining students remain neutral about this fact, where 0 % vote against StudentPIMS. This gives us a good guess that StudentPIMS has the potential to better serve the needs of the students.

TABLE III
STUDENTS WHO ARE SATISFIED WITH EXISTING PIM SOLUTIONS

You are satisfied with existing PIM solutions					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly Disagree	8	11.4	11.4	11.4
	Disagree	28	40.0	40.0	51.4
	Neutral	15	21.4	21.4	72.9
	Agree	17	24.3	24.3	97.1
	Strongly Agree	2	2.9	2.9	100.0
Total		70	100.0	100.0	

TABLE IV
STUDENTS WHO ARE SATISFIED WITH STUDENTPIMS

You are satisfied with StudentPIMS					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Neutral	2	2.9	2.9	2.9
	Agree	21	30.0	30.0	32.9
	Strongly Agree	47	67.1	67.1	100.0
	Total	70	100.0	100.0	

In order to make sure that the difference in satisfaction level between existing solutions and StudentPIMS is significant, we performed paired sample t-test, which enables testing a change in means for a particular variable before and after a particular time, event, and action etc. Here the change in time is the duration for which StudentPIMS was used i.e., three months. For this test the hypothesis and alternative hypothesis are:

H_0 : There is no significant difference between satisfaction levels of students

H_1 : The difference is significant

At 95% confidence interval, we get the results shown in Table V, where the p-value is equal to $0.000 < 0.05$, which rejects the null hypothesis and therefore, we keep the alternative hypothesis valid, which shows that students are much more satisfied with StudentPIMS as compared to existing PIM solutions.

TABLE V
PAIRED SAMPLE T-TEST OF STUDENTPIMS AGAINST EXISTING PIM SOLUTIONS

Paired Samples Test									
		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	You are satisfied with existing PIM solutions - You are satisfied with StudentPIMS	-1.971	1.274	.152	-2.275	-1.668	-12.950	69	.000

When students were asked about the simplicity and usability of StudentPIMS, about 96% (75% strongly agree and 21% agree) responded in favor of the app, whereas only 2.9 % remained neutral. This is shown in Table VI, where none of the registered students answered against the app. When students were asked about whether StudentPIMS respect the privacy of personal information, about 98.6% (70% strongly agree and 28.6% agree) responded in favor of the app, whereas only 1.4% remained neutral. This is shown in Table VII, where none of the registered students answered against the app. Similarly, when they were asked about whether they recommend StudentPIMS for other students to use, all of them responded in favor of this recommendation (Table VIII).

TABLE VI
STATISTICS OF STUDENTS WHO THINK THAT STUDENTPIMS IS SIMPLE AND EASY TO USE

StudentPIMS is simple and easy to use					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Neutral	2	2.9	2.9	2.9
	Agree	15	21.4	21.4	24.3
	Strongly Agree	53	75.7	75.7	100.0
	Total	70	100.0	100.0	

TABLE VII
STATISTICS OF STUDENTS WHO THINK THAT STUDENTPIMS RESPECTS THEIR PRIVACY

You are satisfied with StudentPIMS privacy					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Disagree	1	1.4	1.4	1.4
	Agree	20	28.6	28.6	30.0
	Strongly Agree	49	70.0	70.0	100.0
	Total	70	100.0	100.0	

TABLE VIII
STATISTICS OF STUDENTS WHO RECOMMEND STUDENTPIMS TO STUDENTS

You recommend StudentPIMS to students					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Disagree	1	1.4	1.4	1.4
	Agree	20	28.6	28.6	30.0
	Strongly Agree	49	70.0	70.0	100.0
	Total	70	100.0	100.0	

When students were asked about whether they recommend that StudentPIMS should be made available on Google Play so that other students can download it and use it, about 98.6% (84.3% strongly agree and 14.3% agree) responded in favor of the app, whereas only 1.4% remained neutral. This is shown in Table IX, whereas none of the registered students answered against the app.

TABLE IX
STATISTICS OF STUDENTS WHO RECOMMEND STUDENTPIMS ON GOOGLE PLAY

StudentPIMS should be on Google Play					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Neutral	1	1.4	1.4	1.4
	Agree	10	14.3	14.3	15.7
	Strongly Agree	59	84.3	84.3	100.0
	Total	70	100.0	100.0	

V. CONCLUSION AND FUTURE WORK

The wide spread adoptability and the ubiquitous nature of mobile and smartphones have attracted researchers and developers to utilize its pervasiveness in a number of useful solutions including education. The use of Information & Communication Technologies (ICT) including Smartphones has resulted in fruitful results regarding research (e.g., mobile survey solutions) and education such as distance learning, mobile learning and other useful apps that support learning and training of students. Today, several Android apps are available on Google Play that can be easily, freely downloaded and used on one's Smartphone and tablet.

Among daily life activities of students, curricular and co-curricular activities are much important in order for them to be successful in life and to easily achieve their future endeavors. Students are assigned a number of curricular and co-curricular tasks as part of their education, whose successful completion depends on how much students manage

to remember these tasks in order to be performed in time. Therefore, they use different means such as SMS drafts, mobile phone reminders, sticky notes in printed and digital form, and notes written in their notebooks and on pieces of papers etc., however, due to their limited usability and greater probability of getting lost, these ways of remembering things are limited and involve a great deal of cognitive overload. In order to avoid this, we propose that Smartphones should be used in assisting students in keeping them reminded of what curricular and co-curricular activities that ought to do on their right time, venue, and date. In this regard, several Android-based Smartphone apps are available on Google Play that can be used as PIM solutions; however, these are limited especially in terms of usability, simplicity, privacy and resource-consumption like needing Internet connection and dependency on servers etc. In order to avoid these limitations and to provide students with a more tangible and secure space for managing their personal information regarding curricular and co-curricular activities, we proposed StudentPIMS in this paper and evaluated it through a simple evaluation framework as well as by conducting a survey among the participants of the survey. Results obtained from this evaluation show that students are more likely to use StudentPIMS in comparison to available means of managing personal information.

Currently StudentPIMS facilitates students in managing their personal information regarding curricular and co-curricular activities. However, in order to fully utilize the app, we are planning to include some further features into the app so that not only students but also teachers and researchers take full advantage of the app. The following are some of the recommendations that could be included into the app as future work:

- Modifying the app so that besides students the app can also serve teachers and researchers in doing their activities regarding education and research. The app at startup will ask the user whether they are students, researchers, or teachers and subsequently adapt its functionality accordingly.
- Reporting statistics in the form of a single report that informs the students about completed tasks as well as all pending tasks being categorized into assignments, home works, tests and other co-curricular activities.
- Developing a visualized clustering interface that can present all the activities in the form of clusters so that users can easily browse activities by simply tapping a cluster of interest.

After implementing the app to its fullest by adding these and other possible enhancements, we will upload the app to Google Play as free app so that students, teachers, and most importantly researchers all around the world can take advantage of its services free of cost and to obtain feedback from its users to further improve the app in functionality. We hope that the app will better serve the students in assisting them in managing their personal information.

REFERENCES

- [1] W. Jones. *Keeping Found Things Found: The Study and Practice of Personal Information Management*. Morgan Kaufmann Publishers. ISBN 978-0-12-370866-3, 2008.
- [2] Nouman (January 27, 2015). That You Should Keep in Mind Before Attending Personal Information Management. *Self-Management Skills*. [Online]. Available: <http://www.onionus.com/that-you-should-keep-in-mind-before-attending-personal-information-management/>
- [3] W. Jones. *Personal Information Management*. J. William, Teevan, J. Eds. University of Washington Press, Seattle and London, 2007.
- [4] D. Mizrahi, "Individuality and Diversity among Undergraduates' Academic Information Behaviors: An Explorative Study." *International Journal of Knowledge Content Development & Technology*, vol.3, no.2, pp.29-42, 2013
- [5] M. J. Hashemzadeh and Z. Salehnejad. "Personal information management behavior among postgraduate students of University of Birjand." *2015 4th IEEE Int. Symp. on Emerging Trends and Technologies in Libraries and Information Services*, Noida, January 2015, pp. 307-311
- [6] J. L. Boyles, A. Smith and M. Madden. "Privacy and Data Management on Mobile Devices," Pew Research Center. Internet & American Life Project Report. September 5, 2012, [Online]. Available: <http://pewinternet.org/Reports/2012/Mobile-Privacy.aspx>
- [7] M. Bernstein, M. V. Kleek, D. Karger and M. C. Schraefel. "Information scraps: How and why information eludes our personal information management tools." *ACM Trans. Info. Sys.*, vol. 26, no.4, pp.1-46, 2008

- [8] L. R. Kearns, B. A. Frey, C. Tomer and S. Alman. "A Study of Personal Information Management Strategies for Online Faculty." *Online Learning Journal*. vol. 18, no. 1, 2014, pp. 1-17s
- [9] Power School Mobile. [Online]. Available: <https://play.google.com/store/apps/details?id=com.powerschool.portal&hl=en>
- [10] My Class Schedule: Timetable [Online] Available: <https://play.google.com/store/apps/details?id=de.rakuun.MyClassSchedule.free&hl=en>
- [11] Students: Timetable [Online] Available: <https://play.google.com/store/apps/details?id=crazy.students.student>
- [12] Student Agenda [Online] Available: <https://play.google.com/store/apps/details?id=com.clawdyvan.agendadigitalaluno&hl=en>
- [13] Student Buddy [Online] Available: <https://play.google.com/store/apps/details?id=com.easysolutions.attendancecalculator&hl=en>
- [14] My Study Life [Online] Available: <https://play.google.com/store/apps/details?id=com.virblue.mystudylife>

Contribution to a proportional sharing out of IaaS resources for service adaptation with users services according its profile in cloud computing (An equity based approach)

KANGA Koffi, *Ecole Doctorale Polytechnique de l'Institut Nationale Polytechnique Félix Houphouët Boigny (EDP/INPHB), Côte D'ivoire UMRI 78 Laboratoire de recherche en informatique et télécommunication*

GOORE Bi Tra, *Institut Nationale Polytechnique Félix Houphouët Boigny (INPHB), Côte D'ivoire Laboratoire de Mathématiques et des Nouvelles Technologies de l'Information*

BABRI Michel, *Institut Nationale Polytechnique Félix Houphouët Boigny (EDP/INPHB), Côte D'ivoire Laboratoire de Mathématiques et des Nouvelles Technologies de l'Information*

Abstract-Cloud computing ensures the allowance of resources consumption to the user, by paying for it as he will do for other basic services as water and electricity. In this article we propose an IaaS resource adaptation technique (space capacity) necessary for the SaaS and PaaS in order to improve their functioning in terms of storage capacity by taking into account users' profile. In that way, a proportionality coefficient has been defined, and used for this adjustment and also by taking into account previous IaaS space occupations proportion for each service of cloud. Our contribution is based on the setting up of an allocation technique supported by an algorithm allowing its achievement. The outcome results of the implementation of the algorithm show that our method allows a propositional sharing out of the resources. Therefore the IaaS space should be adapted to the users' service.

Keywords: Cloud computing, Users profile, resources allocation, IaaS resources adaptation.

I. INTRODUCTION

Talking about proportional IaaS resources sharing out between users services based on equity deserves some explanations. In fact, cloud computing is a concept allowing the allocation of resources to the user by paying as he will do for water; electricity and gas. This economic model involves two types of actors namely cloud service provider and the client, the final consumer.

In this model a minimum service is provided free of charge by default over a certain period. This set of services (SaaS, PaaS) cover a certain size of IaaS resources hence (CPU, RAM, DISK SPACE...).

Due to the increasing need of users of cloud services, the IaaS services occupied by them may reach a point of saturation, and therefore might not be able to support the consumption of users. In this regard, user is sometimes tempted to modify or seek on request or by reservation [3] of additional services.

There, arises a need for IaaS allocation resources to the available IaaS or PaaS new list for the user's profile from a given profile.

When the allocation is already made, there have to share out the resource to the different services used by the consumer. As the concept of equity it was developed in the 17th century by Pascal and Ferma [10] further to fairly sharing out of

resources raised by the “Knight of Méré”. That problem was known as the "Problem of portions". It was concerned with the fairly sharing out of the amount bet on during a dice game.

So, Pascal and Fermat tried to find out a method which could allow them to make a fairly sharing out of an amount during a game, if some of the players decide to withdraw before the end. In this case study, there lies the same problem, that is to say how to share out the IaaS resources acquired from a provider among the different user profile and by then between the cloud services.

A. What is means by fairly sharing out?

According to the French dictionary of Emile le Littré (1863), equity is the attitude of giving to each one an equal share, or better to recognize impartiality the right of each one. From this definition, the characteristics of impartiality appears. In the first place it permits not to favor a service at the detrimental of another one, also it must take into account all the users profiles. The second characteristic that stems from the definition is the acknowledgment of each one's rights, in our case all the services and the profiles. In that context, the fairly sharing out would be an operation which would allow all the users of cloud services to benefit from fractions of resources in an impartial manner according to the user's profile.

B. The user's profile

According to [11] the user's profile is part of the different elements of the user's context. This context is defined by Dey in 2000, as a set of elements of information that can be used for characterizing an entity, personal, location or object including users and applications seen as pertinent for the interaction between the users, an application

In cloud computing, the use or the consumption of services would include the following items of the context:

- User profile including statistical data information such as (name, surnames) and evolutive data like localization, time) and the user's preferences.
- Material based resources (screen size types of material, CPU, RAM, bandwidth) operating system version language.
- Session profile represented by connection information (duration, date).

For the sake of contributing to put at the disposal of users ' IaaS resources by taking into consideration his profile in a fairly manner , we propose in this article a propositional mechanism for the allocation of IaaS resources which take into account the scope of each SaaS or PaaS.

For that reason we are defining proportionality based fairly sharing out coefficient. This coefficient is obtained and based on previous use of IaaS by different services.

This paper is organized as follows:

- In section 2 we will be presenting the state of the art, so as to do a review of literature of existing works which seem outstanding.
- Section 3 is concerned with the problem and the purpose of the article
- Section 4 Our proposal
- Section 5 is devoted to the presentation of experimental results

We should end by section 6, with the conclusion and the perspectives.

II. THE STATE OF THE ART

In the search of solution for the allocation of resources in cloud, several works were conducted. Among them they were:

A. User-centered cloud service adaptation: an adaptation framework for cloud services to enhance user experience. International Journal of Computer Integrated Manufacturing [1]

Ubiquitous system and cloud computing have given ways to the development of smart systems able to provide auto-adaptable application allowing the improvement of the quality of the service as well as the experience of the user, according to users reaction and the sensitivity of the context. However, the current approaches of software adaptation have some critical limits when they are applied on software in cloud environments. In fact, according to [1] engineers, of cloud service providers cannot identify users as well as their needs as for their services and its evolution over the time. For that a user centered framework has been proposed, it deals with self-adaptation of cloud computing services according to users need and. Its aim is to see the users of cloud as collaborator in order to provide for their services and needs. From this collaboration the authors have used the interaction between man and machine, existing solutions to collect information on the user of cloud computing.

This collection of data takes into account adaptive services providing, contextually relevant for the task of the users, their behavior and profile. Their services aim at improving the quality of the service, the experience in the use of applications hosted by cloud production systems as a future platform based upon basic model

Limits: Despite the fact that user's profile and behaviors were taken into account, the presented formwork display some limits. In fact, cloud services being hosted in IaaS spaces, it is appropriate to analyze the effect induced by the self-adaptation of those services in the IaaS spaces during the housing. In other words it would be necessary to analyze the side effect of the possible overrunning of the allocated space for each service upon the use of these services and mainly the satisfaction of users.

B. Algorithm based task study planning in a cloud computing environment [2].

This work is a study bases upon task planning methods in cloud computing environment. Here, the authors have stated that the planning task is a NP hard problem, with the growing number of cloud service users and task programming the existing strategies for ordering of tasks can't satisfy the need of users. For that reason, several task planning algorithms have come to being in order to reduce the cloud calculation and also the cost generated by the tasks. In the literature several planning and allocation algorithms to solve this kind of problem have come to being.

In [2] the authors have conducted studies on different algorithms which allow perform the planning and allocation of resources. These are the genetic algorithm, the bees' colony algorithm and the optimization multipurpose swarm particles.

Limits: This work being a study, it was not interested in a work just to contribute in the allocation of resources.

Yet, to provide for solution overrunning to reach the objective of planning and allocation of the resources, he also proposes ways that can lead to find solution concerning problem related to the allocation of the resources. As for the management of users and then profile, it has not been studied by the authors of this work.

C. Cost optimization in dynamic resource allocation using virtual machine for cloud computing environment.[3]

Here, the objective is to reach an optimal solution for the allocation of resources in cloud. For this purpose, the authors purpose two models of resources allocation, booking allocation and request allocation. In fact, the booking allocation

consists for user of cloud to issue a command for the reservation of a service that will be consumed in the future. That is referred to as a long term allocation process. As for the on request based allocation, it consists to ask the service provider to put the resource at your disposal for immediate use; it is a short term allocation. In these two modes, the form of the request, prices of the resources, the latency time and uncertain true consumption factors, by clients are taken into account to adjust compromises between demand and registration. To reach these objectives the authors have used the Bender decomposition algorithm in figure 1 so as to divide the problem of optimization of resource allocation into sub issues, so as to reduce reservation and instantaneous requests cost. The results of these optimization works have shown that resource allocation reservation will be the best to minimize financial cost for users.

Limits: Despite the consideration of financial and time aspects in this form of allocation, technical and those related to volume are not taken into account, also, the user's quality (profile) have not been covered. As for the way in which the allocation is made, the different parameters were included that is to say; time of latency of the resources prices and uncertainty factor to specify the causes of the different request. The causes which could be created by the saturation of spaces ascribed by default by the service provider to the user of the service.

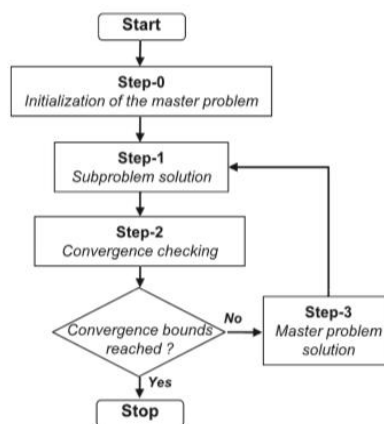


Figure 1 : Diagram of Bender's decomposition algorithm

D. Optimization of resources provisioning cost in cloud computing [5]

In cloud computing, the providers can offer the consumers plans of computer resource provisioning, by reservation or immediate request. Generally the cost of use of allocated computer resources by reservation plan is cheaper compared to the one allocated by the immediate on request plan. Also, the consumer can reduce the total resources provisioning cost. However, the best in advance reservation of resources is difficult to reach owing to the uncertainty of the future resource consumer request and also provider's resource price.

To solve this problem, an optimal cloud resources provisioning algorithm has been proposed based on the "the stochastic programming model". The ORCP (Optimization of Cloud Resources Provisioning) algorithm can allocate long and short term usable quantity of resources.

Limits: Much effort has been made by the authors of this article in searching for solutions to minimize the cost of resource allocation. Effort that is materialized by the combination, the use of diverse algorithms and the results

obtained. Yet, there are limits as for the consideration of aspects related to the user's profile and the effort that the contribution could have on the minimization of sub-provision / over provision of resources. In other terms how will the distribution of resources be made once they are acquired from the providers. The authors then focused on financial aspect of the work.

E. Exploiting dynamic resource allocation for efficient parallel data processing in the cloud. Parallel and Distributed System [6]

In this article, the authors were interested in the use of data in cloud. For that they have discussed challenges and opportunities offered by parallel use of data in cloud computing. From this discussion it appears that the parallel processing of data in cloud can destroy IaaS applications. For it favors the non-utilization of some application hosted in the IaaS. As solution, Nephele has presented as an alternative for a dynamic managing for resource allocation. A sound description of its architecture has been made as well as a comparison of its performances with other solutions such as Map reduce. The appraisal of its performance would give the impression that Nephele is able to (allocate or dis-allocate.) virtual machines for specific task according to its capacity of execution.

Limits: Despite the performances presented by the authors of this framework of which they have spoken. They, also pointed out the limits among which we have the proof of the automatic adaptation incapacity of resources further to an over / under exploitation of the workload

F. A framework for dynamic resource provisioning and adaptation in IaaS clouds. [7]

In cloud computing, IaaS offers opportunity to acquired and allocate computer resources on request so as to achieve a dynamic adaptation of the workload. Here the problem of the dynamic of resources is tackled through the setting up of the supervision mechanism of resources including resources adaptation acquisition selection and a local resources manager compressing task organizer.

In fact, in this work, the resources adaptation algorithm could permit to know the date and the quantity of resources to acquire to release after the occurring of the need for allocation. The algorithm of selection and acquisition of resources could permit to find in a multi-cloud system the appropriate resource at a better price / cost at different providers as for local resource manager from the user. It will load if the selection and acquisition algorithm functions properly, with task organization algorithm.

To bring a solution to these queries, a framework named CREATE endowed with architecture has been set up. It has with in itself different web services among them:

- RSS monitoring and management (RMM) service which permits to manage the configuration and also collect information on the station of the resources and the workload the RSS.
- RSS adaptation service allows the simultaneous management of Several RSS. It keeps up to date a data base of the flow of RSS. It has within itself a set of adaptation algorithm.
- Cloud clustering service (CCS) provides a set of web services of REST type based upon a resource acquisition interface from several service providers...

This set of web service works according to a hybrid adaptation algorithm which permits an optimization of the time of resource allocation for an adaptation (figure 3)

Limits: The aim of this work is to find solution to the issue of a dynamic adaptation of resources. In that way, the authors have set up web services through a hybrid algorithm to optimize the resource provisioning time, a share out cloud context when there is a need for allocation. Despite the interesting results obtained related to the provisioning time, some aspects seem to have been ignored

.These are:

- The distribution of resource with the different cloud providers.
- Consideration of specificities related to the user (profile, context)

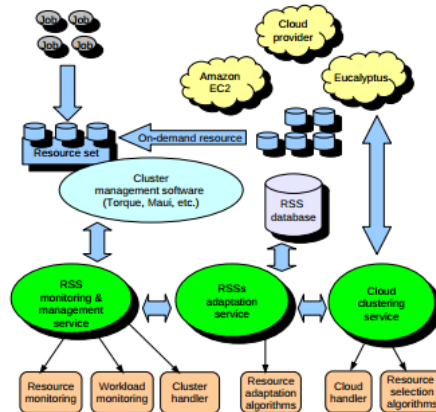


Figure 2 : CREATE Architecture framework

Data: Set of queued, running jobs, resources Q, R, RSS
Result: number of VMs required and to be freed N, F

```

1 initialize  $R_s \leftarrow 0, L \leftarrow 0, W_s \leftarrow 0, S \leftarrow 0, F \leftarrow 0,$   

    $N \leftarrow 0;$ 
2 foreach job  $j \in R$  do
3   if job's remaining walltime  $j_{rt} < 2 \times \delta$  then
4      $R_s = R_s + 1;$ 
5   end
6 end
7 foreach queued job  $j \in Q$  do
8   if job's walltime  $j_{wt} < 2 \times \delta$  then
9      $W_s = W_s + j_{wt};$ 
10  end
11  else
12     $L = L + 1;$ 
13  end
14 end
15 if  $0 < W_s < 2 \times \delta$  then
16    $S \leftarrow 1;$ 
17 end
18 else
19    $S = \frac{W_s}{2 \times \delta};$ 
20 end
21 foreach virtual machine  $v \in RSS$  do
22   if  $v$  has no job then
23      $F = F + 1;$ 
24   end
25 end
26 if  $|Q| > 0$  then
27    $F = 0;$ 
28 end
29  $N \leftarrow (L + S - R_s - F);$ 
30 return  $N, F;$ 

```

Figure 3 : Hybrid and parallel algorithm for service adaptation

G. A toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms [9]

The services hosted under the model of cloud display some allocation, composition, configuration and some complex deployments. In addition, evaluating the performance of provisioning politics, charge application model and political configuration turn out to be difficult.

In this work, the authors propose the use of cloudsim which is an extensible tool allowing to modelize and simulating the cloud computing as so as to overcome challenges.

In addition, the uncertainty constraint related to consumer's request likewise the variation of the prices offered by service providers are considered in the ORCP.

This tool turns out to be compatible with many compatibles with many other operating system. It permits the modelization of the behavior. In fact, in (18) the authors display the different component of cloudsim.

These components are "datacenter", "host" for physical machine, "VM" for virtual ones and cloudlet for mini cloud system. It permits also to implement application provisioning generical techniques that can be extended in comfortable manner and also limit efforts. According to the authors, cloudsim includes the environment of cloud and internetwork connection.

Limits: Given the performance of the tool presented, no IaaS space capacity was taken into account in the different services nor was the user profile, to our knowing. What is more, with regard to the different components of the cloud simulator (cloudsim) information related to preferences a profiles are not included.

III. PROBLEMS

From this review of literature, it appears that the authors have done an excellent work, in the outline of the resources allocation. Yet, this work did not deal with the proportional allocation of resources for a dynamic adaptation to the profile. Here, the profile at stake is the session profile (user premium, gold, bronze) administrator. It comprises the different parameters used by client to log on to the cloud platform. Researchers turn toward the use of framework to carry out the adaptation by resource allocation [1]. [7] Some have conducted research works on the reduction of energy consumption by using resources allocation algorithm forgetting [2] who use the resources allocation to study a colony of bees and a colony of ants.

As for [7], his work was concerned with the optimization of the time of allocation of resources in a multi cloud system in order to realize an adaptation. However as much as I know, none of these works has referred to the resources allocation problem while considering the user's profile to carry out an adaption. Only [7] proposed a hybrid tool permitting the selection and acquisition of resources. In fact, in a previous article, we put into place data model and algorithm allowing to detect user's preferences according to their profile. These profiles use services hosted by IAAS infrastructures. Up on an intensive use of these services on the IAAS, it so happens that those spaces come to saturation and might not be able to support the quantity of data coming from the use of the service. In view of improving the satisfaction of users, the problem of allocation of storage space to different services and in large to different profiles is obvious. Therefore, we propose in this paper, a method in order to favor the allocation of IAAS resources to different user's profile according to the consumption of service and profile. This method will allow us to allocate different resources to the need of consumption for IAAS services (SaaS or PaaS belonging in users' profile).

IV. CONTRIBUTION

Our contribution to the allocation of resources to a cloud computing which consider the profiles of the users consist to propose a data model integrating several parameters. In fact, it includes the following entities:

- User's and their profiles
- SaaS, PaaS and IaaS services consumed by users. This data model will permit us to make up the limit display by the cloudsim simulator by integrating the entities below.

For that reason, we will adapt an approach which will permit us to know the differents services consumed by each profile at two moments:

- At a given date $t = t_0$,
which corresponds to the date of subscription to a profile by the user.
- At any given date corresponding to the date of expression of consumption of additional services.

At this date, our model allows the user to know if he still has sufficient. IaaS resources allowing him to consume additional services. Should the opposite occur our model permits the user either to sacrifice one of the services belonging to a profile, because this service does not have an important use rate or either to keep the service as it is and to proceed to the purchase of additional space from the provider. To reach this scenario, we have introduced in our formalization a parameter belonging in this set $\{-1; 1\}$.

A. Methodology

Allocation consisting in distributing a quantity of resources to a certain number of cloud services. We shall evaluate in the first place, available resources and the volume occupied at an initial date (t_0).

This date corresponds to the date to which the user get access to a new profile, so he will consume a set of a services hosted in an infrastructure of initial size. In a second time, we shall try to evaluate the same metrics after a variation of the consumption of the user's service.

From the instant $t = t_0$, we will have

V_{PjSk} = the space volume occupied by the SK service of a profile (I).

V_{PjSt} = the total size of all the service of a profile (PJ). $VPjSt$ is as $k = m$

$$VPjSt = \sum_{k=1}^{k=m} VPjSk \quad (1)$$

With k = the number of services belonging in a profile. (Pj) a profile number and m = the maximum number of service. Also, all S_k services being hosted by an IaaS infrastructure we can then infer the following relationship.

Let Ipj be the default allocated size to the set of services of a profile.

To avoid space overrunning, we have defined a constraint which is as follows.

$$Ip_j > VPjSt \quad (2)$$

(Pj being a given profile and S_k a SaaS or PaaS service). For a given user, we infer from the relationship (1) the total volume occupied by cell the service of cell the profile this volume is noted: $VPST$: it is such that

$$V_{pst} = \sum_{j=1}^n \sum_{k=1}^m VP_{jSk} \quad (3)$$

With n = the maximum number of profile. Also, at the level of IaaS infrastructures we have.

$$I_{total} = \sum_{j=1}^n I_{p_j} \quad (4)$$

From [2] & [4] we can infer that $I_{total} = \sum_{j=1}^n I_{p_j}$ therefore $I_{total} \geq \sum_{j=1}^n \sum_{k=1}^m VP_{jSk} \quad (5)$

From this relationship, we bring out the proposition represented by the space of the services of a given profile in all the allocated to the space all the services of profile. Let α_{jk} be that proportion.

$$\alpha_{jk} = \frac{\text{Volume of the Profile of a Service}}{\text{Total volume of all services of all profiles}} = \frac{VP_{jSk}}{V_{pst}}$$

$$\alpha_{jk} = \frac{VP_{jSk}}{\sum_{j=1}^n \sum_{k=1}^m VP_{jSk}} \quad (6)$$

At any given date ($t \neq t_0$)

In case the default IaaS space allocated to a profile turns out to be insufficient to contain the services it hosted, or if the user expresses a need to use an additional service, therefore, in a cloud computing environment, this user might acquire additional space so as to satisfy his needs. In this condition, we offer a proportional allocation of this space to the set of services of the users profiles at stake, it then leads to a proportional allocation at all levels of that user's profile. From this allocation there emerges a variable proportionality coefficient according to the number of service and also the size of services by profile.

Definition of the additional space

Let's $I'_{P_jS_k}$ be an additional space allocated to S_k a service of a P_j profile. Follow up the selection of an additional or the withdrawal of an existing service. Let 'n' be a parameter permitting to know the origin of the space to allocate 'n' in such that.

Resources allocation principle

Our allocation principle consists to share out a quantity of resource according to a certain proportion. This proportion is determined by the α_{jk} coefficient in the relationship (6)

Additional proportion after allocation

Let β_{jk} be a new proportion of $I'_{P_jS_k}$ of which benefit.

$$\beta_{jk} = \alpha_{jk} * I'_{P_jS_k} \quad \beta_{jk} = \frac{VP_{jSk}}{\sum_{j=1}^n \sum_{k=1}^m VP_{jSk}} * I'_{P_jS_k} \quad (7)$$

Total volume per service after allocation.

After the allocation, each obtains a new quantity of service to use for its functioning. This volume is sum of the old ($V_{P_jS_k}$) and the β_{jk} . Let $V'_{P_jS_k}$ this new quantity it is expressed as follow:

$$V'_{P_jS_k} = \beta_{jk} + VP_{jSk} \quad (8)$$

$$V'_{PjSk} = \frac{VPjSk}{\sum_{j=1}^n \sum_{k=1}^m VPjSk} * I'_{PjSk} + VPjSk \quad (9)$$

TABLE 1:
SUMMARY OF SERVICE DISTRIBUTION ON ONE ALLOCATION BASED PER USER PROFILE

User' profile	P ₁					P _n				
consummated services	P ₁ S ₁	P ₁ S ₂	P ₁ S _{m-1}	P ₁ S _m	P _n S ₁	P _n S ₂	P _n S _{m-1}	P _n S _m
size of the services	V _{P1S1}	V _{P1S2}	V _{P1Sm-1}	V _{P1Sm}	V _{PnS1}	V _{PnS2}	V _{PnSm-1}	V _{PnSm}
proportion α_{jk}	α_{11}	α_{12}	...	α_{1m-1}	α_{1m}	α_{n1}	α_{n2}	α_{nm-1}	α_{nm}
total volume after allocation	V' _{P1S1}	V' _{P1S2}	...	V' _{P1Sm-1}	V' _{P1Sm}	V' _{PnS1}	V' _{PnS2}	V' _{PnSm-1}	V' _{PnSm}
size of infrastrure	I _{P1}					I _{Pn}				

B. Proposal of algorithm

Our algorithm for the management of the proportional allocation takes:

At the input

- The different user's profile (**P_j**)
- The services (**P_jS_k**) for each profile (**P_j**)
- The volume of resource occupied by each service
- The default size allocated to a profile

At the Out put

We have the new volume (**V'_{PjSk}**) for each service **P_jS_k** after allocation. The execution of this algorithm follows eight (8) different steps.

Steps 1: Identifying all profile for each cloud user.

Identifying SaaS or PaaS service for each profile.

Steps 2: Identifying the IaaS space volumes allocated by default for each profile.

Identifying the volume of different services for each profile.

Steps 3: Calculate the (α_{jk}) proportion for each service

$$\alpha_{jk} = \frac{VPjSk}{\sum_{j=1}^n \sum_{k=1}^m VPjSk} \quad (10)$$

with **n**= maximum number of profile and

m= the maximum number of service for profile

Steps 4 : Identify **S'_k** additional services to consume or to withdraw from the default set.

Identify the volumes **V'_{Sk}** of **S'_k**.

Identify the IaaS (**I'_{Pj}**) resources to allocate to the setoff services.

Steps 5: Calculate the proportion of (**I'_{Pj}**) to allocate to **S'_k**.

$$\beta_{jk} = \alpha_{jk} * I'_{PjSk}$$

Calculate the new size (**V'_{PjSk}**) ascribed to the **S_k** service of the **P_j** profile.

$$V'_{PjSk} = \frac{VPjSk}{\sum_{j=1}^n \sum_{k=1}^m VPjSk} * I'_{PjSk} + VPjSk \quad (11)$$

Algorithm : Pseudo code of pproportional allocation of IaaS resources per user's profile

Input : U : set of user

P : set of all user profile

PS : set of services per profile

Pj : the different profile of the users

PjSk : Sk Service of each profile Pj

Vpjsk : the volume of resources occupied by each Sk by default

I₀ : the size allocated by default by a profiles

Output :

V'PjSk : the new volume of each service PjSk after distribution

Begin

For each $U_i \in U$ do

Write ('Choose the number n of the profiles ' U_i)

read(n)

j is integer

// identification of profiles

Vpst is a decimal = 0

For j = 1 to n do

write ('read the number of m default service of the Pj profile, then read the size of the IaaS space (Ij) allocated to Pj')

read (m), read (Ij)

k is integer

VPS is array of decimal

For k = 1 to m do

write ('capture the size of ' $VPjSk$ 'of PjSk)

readln (VPS[k])

End for

// calculating the total amount of space occupied by all services profile

$Vpst = Vpst + VPS[k]$

// calculation of the fraction of each service

For each Sk of Pj $\in S$

$\alpha_{jk} = Vpjsk / Vpst$

```
end for

end for

// identification of additional services and their respective volumes

VPS' is array of decimal

write ('enter the number x of additional services')

read (x)

b is integer

for b = 1 to x

    read (VPS' [b])

    // Update of the Vpst

    Vpst = vpst+VPS' [b]

End for

// recalculate the proportion represented by the size of a service compared to all the services (by default including additional) per user's profile

 $\alpha' = Vp_jSk / Vpst$ 

// Reading and calculation of the size of new services after allocation of IAAS space to all services

Vnouv is decimal

Write ('Read the new IaaS size to be allocated')

read (IaaS')

// calculation of Vnouv

 $Vnouv = \alpha' * IaaS' + VP_jSk$ 

End for

End
```

V. EXPERIMENTAL RESULT AND IMPLEMENTATION

The appraisal of our solution was achieved in the following conditions:

- The work environment that of cloud report which is a simulator of cloud computing, it is an open source tool whose code source is in Java language.
- We also assumed that after a simulation we have used a set of three (3) users having three profiles (3) per user. Those profiles can support a set of ten (10) services by default. Each of the services has IaaS size occupied and variable according to the service.
- We also assumed that after use, the remaining IaaS space for the second profile was not adapted to the whole need of that same user. Therefore, additional resource has to be found so as to allow the execution or consumption of services of this profile to be made in comfortable manner.

The algorithm we have proposed has given results which are registered in table (2) (for $\eta = 1$) and table (3) for ($\eta = -1$). In those two tables, we have (3) user profile (P1, P2, P3) to which are respectively allocated default storage space (V_{P1St} , V_{P2St} and V_{P3St}). The spaces have following sizes (40, 24 and 19 GB). These services are deployed on those spaces.

- P_{1S1} , P_{1S2} and P_{1S3} for v_{p1st} occupying prospectively the different sizes 14, 16 and 10 Gb
- P_{2S1} , P_{2S2} and P_{2S3} for v_{p2st} occupying prospectively the different sizes 8, 5 et 11 Gb
- P_{3S1} , P_{3S2} and P_{3S3} for v_{p3st} occupying prospectively the different sizes de 6, 9 et 4 Gb

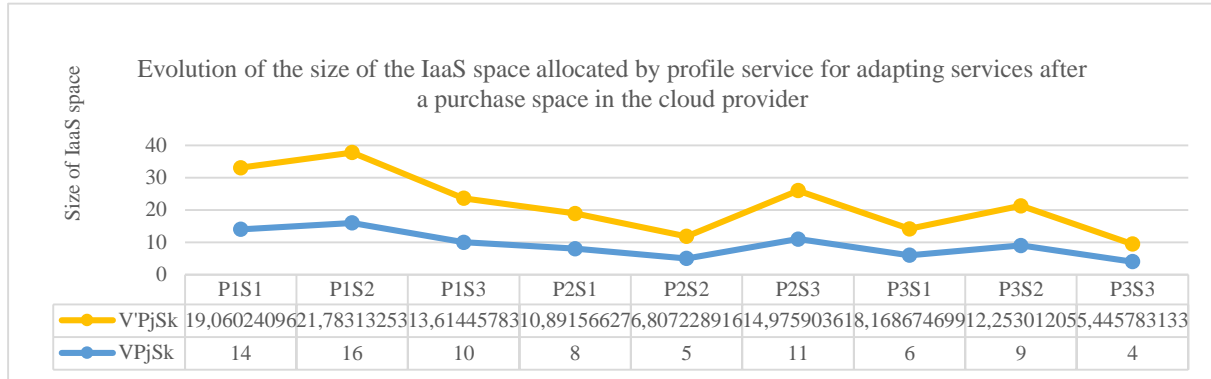


Figure 4 : State of spaces allocated by service allowance $n = 1$

TABLE 2
EXPERIMENTAL DATA FOR $\eta = 1$

User's profil (P_i)	P1			P2			P3			Total
Service (P_{jSk})	P_{1S1}	P_{1S2}	P_{1S3}	P_{2S1}	P_{2S2}	P_{2S3}	P_{3S1}	P_{3S2}	P_{3S3}	9
V_{PjSk}	14	16	10	8	5	11	6	9	4	83
V_{PjSt}	40			24			19			83
α_{jk}	0,168674 7	0,1927710 8	0,1204819 3	0,0963855 4	0,0602409 6	0,1325301 2	0,0722891 6	0,1084337 3	0,0481927 7	1
V'_{PjSk}	19,06024 1	21,783132 5	13,614457 8	10,891566 3	6,8072289 2	14,975903 6	8,1686747	12,253012	5,4457831 3	113
Initial I_{pj}	50			30			20			100
Additional space to allocate I'	30									
Final I'_{pj}	104,4578313			62,6746988			45,86746988			213

B. Interpretation

On the table 2, there are data which correspond to an allocation of resources with the purchase of additional space ($\eta = 1$). Here the user has noticed or been informed by his cloud service provider that his IaaS volume would come to saturation, he then decides to purchase additional space (additional space I') of 3Gb. This space is shared out among the services of the three profiles (P1, P2 and P3). The sharing out is made according to the proportion of occupation of the IaaS (α_{jk}). These proportion being even, they permit the increase of the different volume of IaaS allocated to the services of the different profiles. On figure 4 the curve in blue represents the level of occupied volume after the sharing out of the 30G. As for the one in Brown colour, it shows the level of IaaS after the sharing out of the 30 GB purchased.

The appraisals and the representation of these data through figure 4, leads us to say that the service of the profile at stake have an improve adaptation of this user.

TABLE 3
EXPERIMENTAL DATA FOR $\eta = -1$

User's profil	P1			P2			P3			Total
Service (P _{jSk})	P _{1S1}	P _{1S2}	P _{1S3}	P _{2S1}	P _{2S2}	P _{2S3}	P _{3S1}	P _{3S2}	P _{3S3}	9
V _{PjSk}	14	16	10	8	5	0	6	9	4	83
V _{PjSt}	40			13			19			72
α _{jk}	0,19444444	0,22222222	0,13888889	0,11111111	0,06756757	0	0,08333333	0,125	0,05555556	0,99812312
V' _{PjSk}	16,1388889	18,4444444	11,5277778	9,22222222	5,74324324	0	6,91666667	10,375	4,61111111	93,9793544
Ipj initial	50			19			20			89
Additional space to allocate I'	11									
Final Ipj	96,11111111			34,0952381			41,90277778			171,979354

In table3 above, the data are obtained further to a withdrawal of service ($\eta = -1$). For our experimentation we assume that the users have chosen to delete service P_{2S3} for either financial reason (not purchasing the additional space) or he believes that very is no more use to him. In this conditions the volume of this service in the new configuration will be nil. The 11 GB of space will be shared out to the other services by our algorithm on figure (5) we have presented the level of the different IaaS services through the graphics.

The blue colour curve actually represents the level before the sharing out.

The curve in yellow colour represents the level after the sharing out.

On the graphics, the different volume of service P_{2S3} are marked by zero because that service is not no longer part of the catalogue of services.

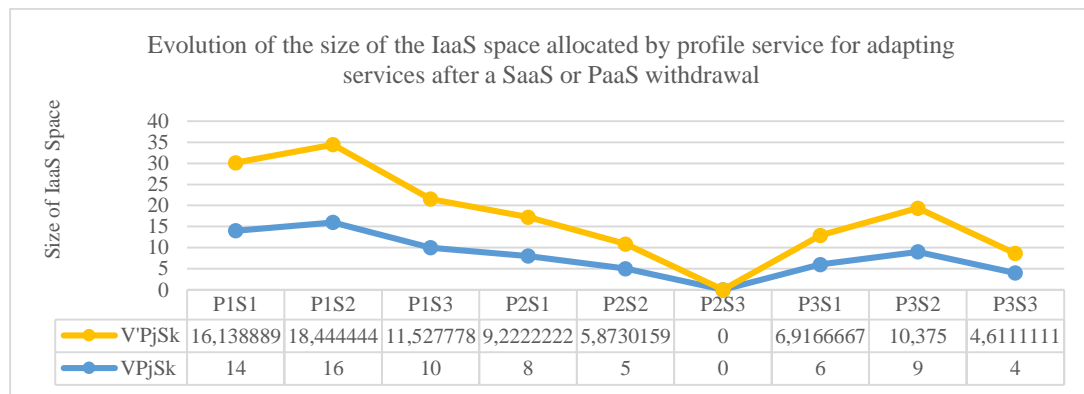


Figure 5 : State of the spaces allocated by service allowance $\eta = -1$

VI. CONCLUSION

The aim of our work consisted in putting into place a method permitting to adapt the volume of IaaS allocated by the service provider to the user to PaaS or SaaS by considering users' profile. For that purpose we have proposed an algorithm allowing a proportional or fairly sharing out of the IaaS additional acquired resources after the purchase from the provider of withdrawal from existing services belonging to several profile users

The analysis of the results show that our methods permit to increase the IaaS resources whatever the provision mode of the resources (additional purchase ($\eta = 1$) or withdrawal of service ($\eta = -1$)) so as to adapt it to the set of service of different profiles in fairly manner.

The satisfaction of users through the adaptation of cloud services being our purpose, we shall orient our next work towards the study of the quality of adjustment of services by considering the different users' profile.

REFERENCES

- [1] Tzafilkou, K., Protogeros, N., & Koumpis, A. (2015). User-centred cloud service adaptation: an adaptation framework for cloud services to enhance user experience. *International Journal of Computer Integrated Manufacturing*, (ahead-of-print), 1-11.
- [2] Bilgaiyan, S., Sagnika, S., Mishra, S., & Das, M. (2015). Study of Task Scheduling in Cloud Computing Environment Using Soft Computing Algorithms. *International Journal of Modern Education and Computer Science (IJMECS)*, 7(3), 32.
- [3] Umamageswari, S., & Babu, M. C. (2014). Cost optimization in dynamic resource allocation using virtual machines for cloud computing environment. *Asia Pacific Journal of Research Vol: I Issue XI*.
- [4] Sato, K., Samejima, M., & Komoda, N. (2013, July). Dynamic optimization of virtual machine placement by resource usage prediction. In *Industrial Informatics (INDIN)*, 2013 11th IEEE International Conference on (pp. 86-91). IEEE.
- [5] Chaisiri, S., Lee, B. S., & Niyato, D. (2012). Optimization of resource provisioning cost in cloud computing. *Services Computing, IEEE Transactions on*, 5(2), 164-177.
- [6] Warneke, D., & Kao, O. (2011). Exploiting dynamic resource allocation for efficient parallel data processing in the cloud. *Parallel and Distributed Systems, IEEE Transactions on*, 22(6), 985-997.
- [7] Duong, T. N. B., Li, X., & Goh, R. S. M. (2011, November). A framework for dynamic resource provisioning and adaptation in IaaS clouds. In *Cloud Computing Technology and Science (CloudCom)*, 2011 IEEE Third International Conference on (pp. 312-319). IEEE.
- [8] Stillwell, M., Schanzenbach, D., Vivien, F., & Casanova, H. (2010). Resource allocation algorithms for virtualized service hosting platforms. *Journal of Parallel and distributed Computing*, 70(9), 962-974.
- [9] Calheiros, R. N., Ranjan, R., Beloglazov, A., De Rose, C. A., & Buyya, R. (2011). CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. *Software: Practice and Experience*, 41(1), 23-50.
- [10] Emmanuel Medernach. Allocation de ressources et ordonnancement multi-utilisateurs : une approche basée sur l'équité. *Operations Research. Université Blaise Pascal - Clermont-Ferrand II*, 2011. French. <Tel-00686891> [Lan et al. 2010]
- [11] Chaari, T., Laforest, F., & Flory, A. (2005, May). Adaptation des applications au contexte en utilisant les services web. In *Proceedings of the 2nd French-speaking conference on Mobility and ubiquity computing* (pp. 111-118). ACM.
- [12] Medernach, E. (2011). Allocation de ressources et ordonnancement multi-utilisateurs: une approche basée sur l'équité (Doctoral dissertation, Université Blaise Pascal-Clermont-Ferrand II).
- [13] Ray, S., & De Sarkar, A. (2012). Execution analysis of load balancing algorithms in cloud computing environment. *International Journal on Cloud Computing: Services and Architecture (IJCCSA)*, 2(5), 1-13.
- [14] KANGA, K., GOORE, B. T., BABRI, M., & OUMTANAGA, S. (2015) Detection of Preference and Selection of Cloud Services for Dynamic Adaptation Based on the User Profile

Enhanced Data Security in Network Environment

Ram Krishna Akuli
Scholar CVRU, Bilaspur,

Dr. J. Durga Prasad Rao
Additional Director and Head,
Computer Science Department
Shri Shankaracharya Mahavidyalaya
Junwani, Bhilai

Dr. Satyendra Kurariya
Head, Computer Science Department
Mata Gujari College
Jabalpur

Abstract - This study is based on the development of a new secure protocol for remote calls. The secure protocol design specification and descriptions are analysed comparing them with the existing protocols. The protocol is designed in a simple way with in built security features. Cryptographic modules can be exchanged due to the flexibility of the new approach depending on various issues and security matters. The developed protocol in this study is platform independent. The security levels of the new secure protocol are properly analysed with desired results. Comparisons with other existing technologies like CORBA or the RMI were also addressed. The results show that creation of a secure network protocol universally acceptable. Although all the bugs and security issues were not addressed as they keep evolving on a daily basis.

Keywords: - *Cryptographic Protocol, Secure Remote Protocol, Network Security*

I. INTRODUCTION

In the modern era that we are living in secure programming, systems have become more and more important. This is as a result of the reliance of computer programs in almost every aspect of our daily activities [1,9,15]. Various achievements have been made to secure programming. For instance, a while ago it was not an easy task to create a connection by use of network and make use of it in programming. The very first network software's, utilized network datagrams and network sockets. These programs were complicated, regular cases of damaged packets and the loss of connection. They were also costly and needed binary data conversion and representation that the programmer would create. Therefore, the need for a cheaper single program that was divided into independent and few sections run and controlled by separate hosts. The available technologies that serve this purpose are RMI, RPC and also CORBA [3,5,17]. These technologies have required engineers to create a program that is similar to that of a single machine. They are less costly for bigger projects and the responsibility for the network layer is transferred from software engine to the network library purveyor.

CORBA, RPC, and RMI technologies are classified as binary protocols that require the programmer to use suitable libraries [7,16,24]. Just like SOAP and REST, they fall in the text-based remote technologies. It should also be noted that these technologies are different in some way. All of them make network communication easier and simple. They permit the programmer to divide the software into sections that are operating at the same time on separate hosts within the network.

Unfortunately, all these technologies have security concerns. Not even one of them is security oriented. It is possible to hack all these remote protocols which can be dangerous in different situations. The security mechanisms that are applied in these remote protocols are not safe for example, SSL, SSH, Kerberos and many others. The vulnerabilities and weak authentication in the process of authorization cannot rely on cryptographic algorithms alone. In the case of software libraries that were established with programming guidelines that were secure are also still not safe. Attackers can make good use of security bugs such as incorrect permission, poor authentication, buffer overflows and many others [7,13,14]. This created the need to create a network protocol that possesses security mechanisms that deal appropriately with the current requirements. Therefore, this study shows that there is potential for creating design, various specification and implementation of a universal, new network protocols intended for remote function, procedures or even methods that is characterized by:

- a. Security mechanisms that are build-in
- b. The security mechanisms will be suited in a manner that aligns the security level to the demand of various systems utilizing the protocol. This will involve even the legal requirements of the client.
- c. The new protocol will have no operating system dependencies and hardware.

The Secure Remote Protocol (SRP) addressed in this study fits the characteristics mentioned above.

II. SRP SPECIFICATIONS

Level of architecture in SRP

The aim of this study is to come up with a simple and easy protocol, basing itself on simple architecture. The Secure Remote Protocol (SRP) is designed to be independent of the programming language that is used and system of architecture [2,18,28]. It comprises of some stubs that are allocated to various languages and platforms. It is from the secure interface definition language that the software developer should produce the source code and instruct the functions existing in produced code [8,21,26]. The produced codes should utilize stubs to start the remote methods. This applies to other remote protocols like COBRA, RMI or RCPC [1,4,25,27]. It should also be noted that in SRP stubs are in control of authorization, communication, and network communication.

The programmer must be aware of the signature of the used remote methods. He is not required to be aware of the IP address or port. The main registry, stores port, IP address and other data are associated with the method. The local cache is added to the whole architecture. Cryptographic modules control decryption, encryption, together with hash functions that are responsible for data integrity and confidentiality [6,7,22,23]. This means that developers were not required to carry the cryptography about [37]. Another important component that is available in the SRP is the 'Authentication and Authorization Manager' better known as the AA manager. Authenticated users were obliged to conduct all the revocations. Furthermore, each and every remote action ought to be permitted. Developers are also allowed to implement the AA Manager. It is very simple architecture as programmers produce sources from files of

interface description and later form their program [34,35,36]. The other components were conveyed by the developer of the SRP library.

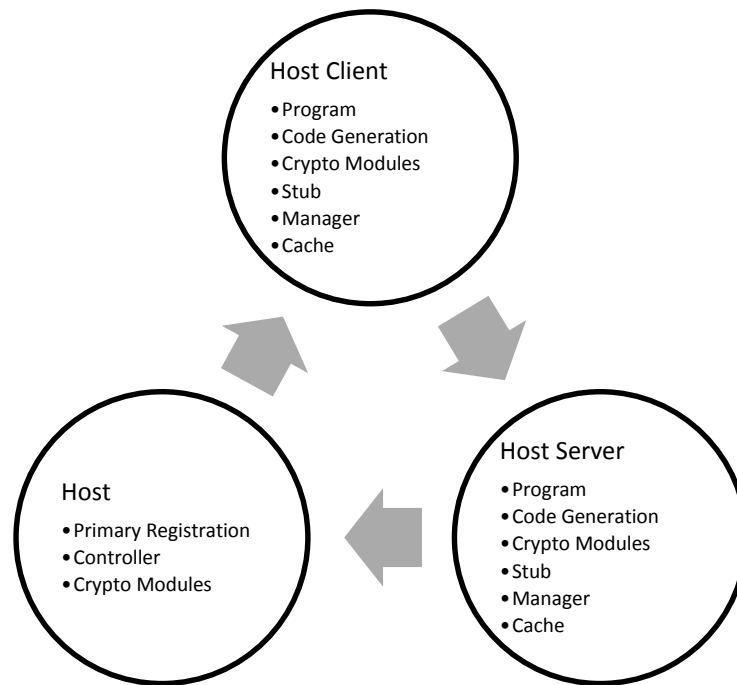


Figure 1: Simple architecture of SRP

III. SRP DATA REPRESENTATION

Since the protocol is so much technical and detailed, data representation ought to be appropriately defined. Simple types and basic data were sent to the big endian representation, where they are converted by stubs into local architecture if they are not compatible with the architecture available [38,39]. The whole process was carried out transparently. Each type was numbered with a unique identification number to differentiate it from the rest. The identification numbers are usually required by the protocols and are useful in data representation [3,11,12,29]. Data types that were used are Boolean, string 8, string 16, int16, int64, int32, int8, string 24 and many more.

Network

The frames of the network that were sent via the network connecting to hosts contained two sections of the frame. The first section was a header, and the other was the encrypted part. [10,31,32] In the SRP, the client had the mandate to keep the encrypted part confidential or he or she could share it.

IV. SECURE INTERFACE DESCRIPTION LANGUAGE (SIDL)

Secure Interface Description Language (SIDL) also forms the specifications of SRP the language bases itself on XML. Each of the specifications that form the SIDL is named to be identified by the code generator. The names given are useful in file names and classes. Every function(C), methods (the language of the object) and remote procedure (Pascal) are referred as function. All the functions had different returned type and name [33,34,35]. All the parameter had two mandatory XML parameters: type and name. The produced codes had a total of three parts: client, server and common. The common part was supposed to be used by both the server and the client. It controlled the process of registration of functions that was performed in the Main Registry. In the part client functions that were ready to be used were present. In the SRP, the remote method implementation is the duty of the programmer.

SRP was divided into Java sections such as `utils.crypt`, `connectors`, `utils.packets`, `utils.types` generated code or components. In each divided section, there were groups of the same class. For instance, the `utils` type was responsible for handling all the SRP types, the `connectors` had classes that controlled connections involving classes from various components like the Stub or Main Registry. Every connection was responsible for any exception. These exceptions were like receiving inaccurate data, tracking lost connections or losing connection [40,41]. The mechanisms of Cryptographic in SRP are build-in in the `utils.crypt` packages adapters [2,19,30]. The classes offer support to appropriate algorithms. The developer of the software, in this case, is not required to carry out the process of proper implementation. In the issue of authentication and authorization, the software engineer was given the mandate to control it. The default mechanisms also are build-in although if changes are necessary, they were allowed.

SIDL code generator was written to form a complete implementation of the SRP. This part was not required by the user but was meant for the software developer to produce stubs from the files that had interface description in their SIDL. In this case, the parser language is independent and universal. The output, on the other hand, was Java code. However, changes can be done at any time.

Quality of the code

To come up with a secure programming network as intended, the codes were protected by a series of test cases. These tests were: integration tests, unit tests, sanity tests and component tests cases [3,42,43]. Network scanners were never utilized because there was an absence of attack vectors linking with SRP. All the bugs that we are all familiar with were fixed in place as implementation was being carried out. The source codes were inspected, and the necessary security issues were addressed appropriately. The find bugs were able to locate the vulnerabilities. It was also realized that the find bugs software could not function if the 'white character' were present in the route leading to the executable file.

V. IMPLEMENTATION OF SRP

Cryptography-related stuffs

Cryptography was necessary because clear-text protocols are never secure because they can be breached or faked at any time. Cryptographic algorithms provide confidentiality messages over the network that is sensitive and classified. Additionally, personal data is similarly protected. Mechanisms of data integrity also were implemented by use of the one-way hash functions. Cryptographic mechanisms also provided the protection from replay attacks with build-in mechanisms into the protocols [44,45,46]. Due to the dangerous state of the replay attacks, the random digits that are usually used to protect protocols were randomized in the most complex way. Software that is secure with the quality random generator was put in place. In general, the SRP fulfilled all the requirements of cryptography mechanisms.

Authentication, Authorization, and Accounting

The process of authentication involved identifying and verifying the user who wishes to access or use the device. The whole process creates a trustful relation between the application and the user. The authentication procedure depended on the type because there are many types of authentication available. The SRP provided a mutual authentication mechanism that is the safest way of authentication [47,48]. The both sides involved had to authenticate each other. There were few types of authentication mechanism. For instance, the first type is involved pin codes or passwords. The passwords rules stated that they are not to be shared, as they are used as identities of the users. It is advisable that we use at least two types of authentication in our networks to protect it efficiently. The reason behind the use of more authentications is to make it hard for intruders to fake or breach them.

Since authentication is not sufficient alone, authorization was also used. This is the processes of investigating the user's rights. This gave the need for proper user management and appropriate configuration in the SRP. The privileges of the users were limited to prevent them from conducting unimportant actions. Therefore, rule-based access control (RBAC), Mandatory access control (MAC), role-based access control (RBAC) and discretionary access control (DAC) were all put in place [2,49, 50]. In SRP, the software designer is allowed to use authorization that is a default. The designer can also be in full control and offer authorization to other parts depending on the system requirement.

On the other hand, accounting was necessary, which is the presence of logs originating from operations that have been performed. Only authorized and authenticated users were allowed to carry out accounting.

Denial-of-service

To offer full protection, the SRP had to be able to be secure from Denial-of-service (Dos) attacks. To come up with such high-quality protection, the protocol had to possess limits for specific operations [51,52,53]. The administrator was responsible for such limits and controlled them. The origin of the attacks was also supposed to be blocked. The issue of attack was left open in the implementation of the SRP.

Input validation

Most threats occur as a result of input validation that is insufficient. To secure the SRP, the input specified by the protocol description should be validated appropriately. Additionally, every program was supposed to validate each input despite the fact that the protocol implementation is protected. SRP, which is implemented in Java, deals with input validation as expected [2]. Race conditions issues were also addressed during the designing period, with the protocol serializing all the network requests.

Third-party software and unsafe functions

It is sometimes hard to determine whether the third-party software's that were utilized in the implementation of protocols are free from vulnerabilities and bugs [54,55,56]. Therefore, the software licenses and legal issues were well addressed to prevent risks that end-users and developers take. The SRP licenses were present and valid and were utilized with Java Runtime Environment licenses.

Static Analysis Tools

All the warnings were addressed and fixed in the secure implementation. Static analysis tools like Klocwork and Yasca were utilized [66,67]. These tools diagnose problems that are related to security such as race conditions, memory leaks, unsafe functions usage, insufficient input validations and many more issues. The SRP took such security issues into consideration and were fixed appropriately. For instance, the FindBug was utilized for inspecting Java implementation, which also checked that present codes were safe from attacks.

VI. SECURITY ASSESSMENT AND SECURITY-RELATED TESTS

The SRP was tested to ascertain if it was secure and protected as it was intended. There many methodologies that can be used to test security related matter. Code average metric is common and popularly used metric when it comes to tests [7,57,58]. The entire source codes are supposed to be covered with different types of tests from unit tests to the last one which is the system test. Many tests as possible were supposed to be automated. The importance of security-related tests has become of great importance and in this study, they were not assumed. The security tests performed white-box and black-box penetration tests. If the software source code is not a public one, then it becomes known that the attacker is carrying out tests that are black-box [2,59]. In the case of white-box, the tester is aware and knows what the product being tested is. The importance of the penetration tests was to uncover new or previously known vulnerabilities that exist. Fuzz testing was also to be conducted; that is also a technique used in testing soft wares. Fuzz testing comprises of giving unexpected, random or invalid data straight to the inputs of particular software. The particular software is consequently tested for any failings or crashes of the code assertions that is build-in. The two types of fuzzing tests were utilized: mutation-based and generation-based. Fuzz testing is linked with the white-box and the black-box tests [8].

After the specification and implementation stage of the SRP, the following security assessments were tested and the methods of prevention designed appropriately as represented in the table below.

Potential Attack	Prevention Methods
Shared key	Dedicated key protects important data
Port scanning	Specific host and port for each process.
Sniffing	Configuration and encryption as per administrator requirements.
Unauthorized users	Addressed mechanisms of authorization and authentication.
Spoofing	Encryption of credentials
Replay attack	Unique call number replay
Poisoning of the local cache	Components responses and request were authenticated and encrypted
Service-Denial	Maximum load parameter addition in the configuration
Used algorithms that are weak	Permission was granted for more cryptographic algorithm.
Dictionary attack	Strong keys and passwords were forcefully generated for the protocol designer and implementation.

Table 1: Potential attacks and prevention methods

VII. PERFORMANCE RESULTS

The SRP was tested on a 2410M processor that is i5 regarding an Intel Core. The software was a VMware player. Each virtual machine had individual one core computer processing unit and a memory of 1GB. The operating system that was in use was named Fedora with a Core 15. The results obtained from the performance were in resemblance of those of remote method that was implemented with a mechanism of Java RMI [60,61]. The implementation of CORBA in Java also gives the same results just like this study [8,62]. The following table shows the result of SRP compared to that of CORBA and RMI.

Test	Requests per second
SRP flawless test with no hash	612
SRP flawless test that is MD5	513
SRP flawless text that is SHA1	498
SRP OFB or AES that is SHA1	502
RMI Java	399
CORBA Java	497

Table 2: SRP results compared to CORBA and RMI

VIII. CONCLUSION

According to the finding and the purpose of this study, it is possible to create a Secure Remote protocol. It can be defined with certain specification and implemented appropriately. The SRP should have a simple architecture and components to be able to address security issues more effectively and in a faster way. The SRP composed of an easy, new Secure Interface Description Language (SIDL) that is XML-based. The language is critical when it comes to remote protocols meant for the purpose of the interface definition. The SRP library is also present together with a translator of the SIDL. The quality of the present codes of SRP was assessed. Furthermore, the whole programming practices in use were secure, which means that the SRP implementation was also secure. It should be worth noting that some security testers were absent meaning that there are some doubts, on the whole, implementation process being secure. Security guidelines were also addressed in the study, with various security tests and assessments conducted. The assessment process was conducted, and the methods of prevention were also outlined properly. The performance results were also analysed and compared to those of Java CORBA and Java RMI.

As a result of this study, there are areas that have not been explored or addressed and further development is needed. Each and every day, new bugs and security issues arise. The security issues keep on changing as attackers adapt to the security tests and mechanisms that are present or have been created. The SRP is a very simple architecture and can be modified to suit certain requirements of clients [2,63,64]. Some of the security matters have not been well addressed. This leaves room for further development and innovation on security-related matters. Additionally, implementation by use of native languages may result in faster connections and network processes [7,65]. To conclude, this study shows that there is potential for creating the design, various specification and implementation of a universal, new network protocols intended for remote function, procedures or even methods.

REFERENCES

- [1.] Dong, L., & Chen, K. (2012). *Cryptographic protocol: Security analysis based on trusted freshness*. Berlin: Springer
- [2.] Medhi, D., & Ramasamy, K. (2007). *Network routing: Algorithms, protocols, and architectures*. Amsterdam: Elsevier/Morgan Kaufmann Publishers.
- [3.] Williams, A. (2001). *Java 2 network protocols black book*. Scottsdale, Ariz: Coriolis Group Books.
- [4.] Kizza, J. M. (2009). *A guide to computer network security*. London: Springer.
- [5.] Sutton, M., Greene, A., & Amini, P. (2007). *Fuzzing: Brute force vulnerability discovery*. Upper Saddle River, N.J: Addison-Wesley.
- [6.] Latze, C. (2010). *Towards a secure and user friendly authentication method for public wireless networks*. Berlin: Logos.
- [7.] Erickson, J. (2009). *Database technologies: Concepts, methodologies, tools, and applications*. Hershey, PA: Information Science Reference.
- [8.] Canetti, R. (2001, October). Universally composable security: A new paradigm for cryptographic protocols. In *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on* (pp. 136-145). IEEE.
- [9.] Abadi, M., & Gordon, A. D. (1997, April). A calculus for cryptographic protocols: The spi calculus. In *Proceedings of the 4th ACM conference on Computer and communications security* (pp. 36-47). ACM.
- [10.] Boneh, D., DeMillo, R. A., & Lipton, R. J. (1997, May). On the importance of checking cryptographic protocols for faults. In *Advances in Cryptology—EUROCRYPT'97* (pp. 37-51). Springer Berlin Heidelberg.
- [11.] Canetti, R. (2000). Security and composition of multiparty cryptographic protocols. *Journal of CRYPTOLOGY*, 13(1), 143-202.
- [12.] Abadi, M., & Gordon, A. D. (1998). A bisimulation method for cryptographic protocols. In *Programming Languages and Systems* (pp. 12-26). Springer Berlin Heidelberg.
- [13.] DeMillo, R. A., Lynch, N. A., & Merritt, M. J. (1982, May). Cryptographic protocols. In *Proceedings of the fourteenth annual ACM symposium on Theory of computing* (pp. 383-400). ACM.
- [14.] Mao, W. (2003). *Modern cryptography: theory and practice*. Prentice Hall Professional Technical Reference.
- [15.] Meadows, C. (1996). The NRL protocol analyzer: An overview. *The Journal of Logic Programming*, 26(2), 113-131.
- [16.] Mitchell, J. C., Ramanathan, A., Scedrov, A., & Teague, V. (2006). A probabilistic polynomial-time process calculus for the analysis of cryptographic protocols. *Theoretical Computer Science*, 353(1), 118-164.
- [17.] Comon, H., & Cortier, V. (2005). Tree automata with one memory set constraints and cryptographic protocols. *Theoretical Computer Science*, 331(1), 143-214.
- [18.] Amadio, R. M., Lugiez, D., & Vanackère, V. (2003). On the symbolic reduction of processes with cryptographic functions. *Theoretical Computer Science*, 290(1), 695-740.
- [19.] Comon-Lundh, H., & Cortier, V. (2003, June). New decidability results for fragments of first-order logic and application to cryptographic protocols. In *Rewriting Techniques and Applications* (pp. 148-164). Springer Berlin Heidelberg.
- [20.] Boyd, C. (1990). Hidden assumptions in cryptographic protocols. *IEE Proceedings E (Computers and Digital Techniques)*, 137(6), 433-436.
- [21.] Jorba, A. R., & Roca, J. C. (2007). *U.S. Patent No. 7,260,552*. Washington, DC: U.S. Patent and Trademark Office.
- [22.] Deng-Guo, Z. Y. B. F. (2006). Design and Analysis of Cryptographic Protocols for RFID [J]. *Chinese Journal of Computers*, 4, 008.
- [23.] Gordon, A. D., & Jeffrey, A. (2004). Types and effects for asymmetric cryptographic protocols. *Journal of Computer Security*, 12(3, 4), 435-483.
- [24.] Argyroudis, P. G., Verma, R., Tewari, H., & Mahony, D. O. (2004, August). Performance analysis of cryptographic protocols on handheld devices. In *Network Computing and Applications, 2004.(NCA 2004). Proceedings. Third IEEE International Symposium on* (pp. 169-174). IEEE.
- [25.] Fiore, M., & Abadi, M. (2001, June). Computing symbolic models for verifying cryptographic protocols. In *csfw* (p. 0160). IEEE.
- [26.] Maggi, P., & Sisto, R. (2002). Using SPIN to verify security properties of cryptographic protocols. In *Model Checking Software* (pp. 187-204). Springer Berlin Heidelberg.
- [27.] Blanchet, B., & Podelski, A. (2005). Verification of cryptographic protocols: tagging enforces termination. *Theoretical Computer Science*, 333(1), 67-90.
- [28.] Jorba, A. R., & Roca, J. C. (2007). *U.S. Patent No. 7,260,552*. Washington, DC: U.S. Patent and Trademark Office.
- [29.] Pass, R., & Rosen, A. (2008). New and improved constructions of nonmalleable cryptographic protocols. *SIAM Journal on Computing*, 38(2), 702-752.
- [30.] A. K. Dey, J. D. P. Rao, and T. D. Singh, Energy saving issue in Mobile Ad-hoc Networks. ISBN: 978-81-923288-1-2, 2013.
- [31.] J. D. P. Rao and R. K. Akuli, "A Brief Study on Measures to Improve Cyber Network Security," pp. 20–22, 2015.
- [32.] R. K. Akuli, J. D. P. Rao, and S. Kurariya, "A STUDY OF SECURITY MECHANISMS IMPLEMENTED IN NETWORK PROTOCOLS," Indian Streams Res. J. ISSN 2230-7850, vol. 5, no. 11, pp. 1–3, 2015.
- [33.] R. K. Akuli, J. D. P. Rao, and S. Kurariya, "NETWORK SECURITY MECHANISMS THROUGH OSI/ ISO NETWORK MODEL FOR UPPER LAYERS," Golden Res. Thoughts ISSN 2231-5063, vol. 5, no. 6, pp. 1–4, 2015.
- [34.] Taylor, D., Wu, T., Mavrogiannopoulos, N., & Perrin, T. (2007). Using the Secure Remote Password (SRP) protocol for TLS authentication. *IETF Request for Comments*, 5054.
- [35.] Bheemarasetti, S., & Prathuri, C. (2002). *U.S. Patent Application No. 10/189,058*.
- [36.] Duncan, R. G., & Shabot, M. M. (2000). Secure remote access to a clinical data repository using a wireless personal digital assistant (PDA). In *Proceedings of the AMIA Symposium* (p. 210). American Medical Informatics Association.
- [37.] Kim, S. K., & Chung, M. G. (2009). More secure remote user authentication scheme. *Computer Communications*, 32(6), 1018-1021.
- [38.] Humenberger, M., Schraml, S., Sulzbachner, C., Belbachir, A. N., Srp, A., & Vajda, F. (2012, June). Embedded fall detection with a neural network and bio-inspired stereo vision. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on* (pp. 60-67). IEEE.

- [39.] Abbott, B. P., Abbott, R., Adhikari, R., Ajith, P., Allen, B., Allen, G., ... & Araya, M. (2009). Search for gravitational-wave bursts in the first year of the fifth LIGO science run. *Physical Review D*, 80(10), 102001.
- [40.] Jahnke, J. H., D'entremont, M., & Stier, J. (2002). Facilitating the programming of the smart home. *Wireless Communications, IEEE*, 9(6), 70-76.
- [41.] Garavel, H., Lang, F., & Mateescu, R. (2001). An overview of CADP 2001.
- [42.] Baggen, R., Correia, J. P., Schill, K., & Visser, J. (2012). Standardized code quality benchmarking for improving software maintainability. *Software Quality Journal*, 20(2), 287-307.
- [43.] Boissinot, B., Darté, A., Rastello, F., De Dinechin, B. D., & Guillon, C. (2009, March). Revisiting out-of-SSA translation for correctness, code quality and efficiency. In *Proceedings of the 7th annual IEEE/ACM International Symposium on Code Generation and Optimization* (pp. 114-125). IEEE Computer Society.
- [44.] Lee, W. B., & Lee, C. D. (2008). A cryptographic key management solution for HIPAA privacy/security regulations. *Information Technology in Biomedicine, IEEE Transactions on*, 12(1), 34-41.
- [45.] Bishop, M. A. (2002). The art and science of computer security.
- [46.] Khalili, A., Katz, J., & Arbaugh, W. A. (2003, January). Toward secure key distribution in truly ad-hoc networks. In *Applications and the Internet Workshops, 2003. Proceedings. 2003 Symposium on* (pp. 342-346). IEEE.
- [47.] Glass, S., Hiller, T., Jacobs, S., & Perkins, C. (2000). *Mobile IP authentication, authorization, and accounting requirements* (No. RFC 2977).
- [48.] Aboba, B., & Wood, J. (2003). Authentication, authorization and accounting (AAA) transport profile.
- [49.] Karpati, D., Zilberman, A., Amos, E. B., & Halevi, I. (2006). *U.S. Patent Application No. 11/481,858*.
- [50.] Moustafa, H., Bourdon, G., & Gourhant, Y. (2006, September). Authentication, authorization and accounting (AAA) in hybrid ad hoc hotspot's environments. In *Proceedings of the 4th international workshop on Wireless mobile applications and services on WLAN hotspots* (pp. 37-46). ACM.
- [51.] Ferguson, P. (2000). Network ingress filtering: Defeating denial of service attacks which employ IP source address spoofing.
- [52.] Jung, J., Krishnamurthy, B., & Rabinovich, M. (2002, May). Flash crowds and denial of service attacks: Characterization and implications for CDNs and web sites. In *Proceedings of the 11th international conference on World Wide Web* (pp. 293-304). ACM.
- [53.] Kuzmanovic, A., & Knightly, E. W. (2003, August). Low-rate TCP-targeted denial of service attacks: the shrew vs. the mice and elephants. In *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications* (pp. 75-86). ACM.
- [54.] Qin, F., Wang, C., Li, Z., Kim, H. S., Zhou, Y., & Wu, Y. (2006, December). Lift: A low-overhead practical information flow tracking system for detecting security attacks. In *Microarchitecture, 2006. MICRO-39. 39th Annual IEEE/ACM International Symposium on* (pp. 135-148). IEEE.
- [55.] Giffin, D. B., Levy, A., Stefan, D., Terei, D., Mazières, D., Mitchell, J. C., & Russo, A. (2012). Hails: Protecting data privacy in untrusted web applications. In *Presented as part of the 10th USENIX Symposium on Operating Systems Design and Implementation (OSDI 12)* (pp. 47-60).
- [56.] Shao, Z., Xue, C., Zhuge, Q., Qiu, M., Xiao, B., & Sha, E. H. M. (2006). Security protection and checking for embedded system integration against buffer overflow attacks via hardware/software. *Computers, IEEE Transactions on*, 55(4), 443-453.
- [57.] Subashini, S., & Kavitha, V. (2011). A survey on security issues in service delivery models of cloud computing. *Journal of network and computer applications*, 34(1), 1-11.
- [58.] Curphey, M., & Arawo, R. (2006). Web application security assessment tools. *Security & Privacy, IEEE*, 4(4), 32-41.
- [59.] Chellappa, R. K., & Pavlou, P. A. (2002). Perceived information security, financial liability and consumer trust in electronic commerce transactions. *Logistics Information Management*, 15(5/6), 358-368.
- [60.] Bennani, T., Blain, L., Courtes, L., Fabre, J. C., Killijian, M. O., Marsden, E., & Taiani, F. (2004, June). Implementing simple replication protocols using corba portable interceptors and java serialization. In *Dependable Systems and Networks, 2004 International Conference on* (pp. 549-554). IEEE.
- [61.] Siegel, J. (2000). *CORBA 3 fundamentals and programming* (Vol. 2). New York, NY, USA: John Wiley & Sons.
- [62.] Gray, N. A. (2004, April). Comparison of Web Services, Java-RMI, and CORBA service implementations. In *Proceedings of the 5th Australasian Workshop on Software and System Architectures at ASWEC* (pp. 52-63).
- [63.] Lee, W. J., Shin, Y., Lee, J., Kim, J. W., Nah, J. H., Jung, S., ... & Han, T. D. (2013, July). SGRT: A mobile GPU architecture for real-time ray tracing. In *Proceedings of the 5th high-performance graphics conference* (pp. 109-119). ACM.
- [64.] Blodget, B., James-Roxby, P., Keller, E., McMillan, S., & Sundararajan, P. (2003). A self-reconfiguring platform. In *Field Programmable Logic and Application* (pp. 565-574). Springer Berlin Heidelberg.
- [65.] Almeida, N., Busson, P., Faure, J. L., Gachelin, O., Gras, P., Mandjavidze, I., ... & Varela, J. (2005). The selective read-out processor for the CMS electromagnetic calorimeter. *Nuclear Science, IEEE Transactions on*, 52(3), 772-777.
- [66.] Topuz, B. (2010). Development of a static analysis tool to find security vulnerabilities in java applications.
- [67.] Aner, Y. (2009). *Securing the Sage Notebook* (Doctoral dissertation, Master's thesis, Royal Holloway, University of London).

Security Concerns with Open Research Issues of Present Computer Network

Geetanjali Rathee, Hemraj Saini

Department of Computer Science and Engineering, Jaypee University of Information Technology, Waknaghat,
Solan-173234, Himachal Pradesh, INDIA

Abstract-Present networks are the mainstay of modern communication. The existence of networks is enriching our society in countless different ways. Now days, wireless mesh network is considered as an auspicious technology for posing self-healing, organizing and configurable capabilities but one of the foremost challenge in the enterprise of these networks is their susceptibility to security assaults (eavesdropping, network layer attacks and denial of service). In order to overcome against these assaults, several security anxieties are proposed but authentication is taken as an important parameter to provide a secure communication. In this chapter, a review is discussed from origin to the current networking technology i.e. WMN. In addition to this, WMN security is concerned with recent applications such as smart grids, intelligent transportation system, multimedia systems etc. further a clear overview of security with respect to each layer is elucidated and finally the chapter is ruined by outlining the future work which is the next step of this research

I. INTRODUCTION

The revolution of computers in 1990, now continuing in the 21st century involves “computer networks” [1]. A computer network or simply a network is a collection of computers that allows a computer or device to exchange the data or sharing of hardware and information [2]. Today, networks are the mainstay of modern communication. The existence of computer network is enriching our society in countless different ways.

The birth of computer networks initialized in 1940 when George Stibitz used a teletype machine to send the instruction for a problem set. In 1950, SAGE (Semi-Automatic Ground Environment) military radar systems were used for communicating the network. Further adding to it in 1960, SABRE went online with 2 connected mainframes. In 1962, JCR Licklider [3,4] developed a working group and introduced an interest at the ARPA i.e. Advanced Research Project Agency [4]. In continuation with, a DARPA program was initialized entitled “Resource Sharing Computer Networks” in 1969. The DARPA program had the following objectives: i) to develop techniques on interconnecting computer. ii) To increase the productivity of resource sharing. After that in 1991, home broadband was created. Home broadband enters into mainstream usage and begins growing at a faster rate in 2001. To show the sequence and revenue growth, 10 GE (Giga Ethernet) market was launched in 2001. Infact! Today 100 GE standards are fully completed. Table 1 shows the brief explanation of computer network history.

As the organizations rely heavily on the ability to share information in an efficient and productive manner [4], Computer networks are now the part of almost each and every business in which computers can seek a pathway anyhow [5]. When it comes to setting up a network, an organization has two options; wired network and wireless network. Let us discuss a brief introduction on networking as shown in figure 1.

A. Types of Network

Wired networking is the most common type of LAN technology, which is also called Ethernet network [6]. In this network, the connections among computers or devices are made using a physical wire or cable [7, 8]. It is simply a collection of two or more computers connected through Ethernet cables. To connect a computer or a device to the

network, an Ethernet adapter is used, which connects the devices either internally or externally. The wired network is further divided into two parts; Point to Point and Multipoint.

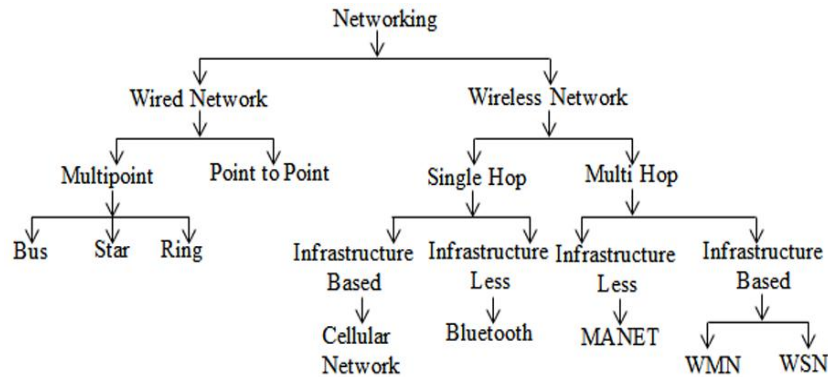


Figure 1 Network Classification

Point to Point network uses an actual length of cable to connect two ends of devices and provides a dedicated link between these two devices [9, 10]. In this network, the entire capacity of the link is reserved only between the two devices as depicted in figure 2 (a). *Multipoint network* is one in which more than two devices share a single link [11] as shown in figure 2(b). There are basically three network topologies in multipoint networking. *Star Network* is generally a naive type of network, which has two or more computers connected to one central hub [12] and this type of network is to be used for small business and home network. Figure 2(c) shows the diagram of star networking [13].

TABLE I
HISTORY OF COMPUTER NETWORKS

Computer Network History		
S.NO.	Time	Description
1	1940	Teletype machine was used to send instruction for a problem set
2	1950	The SAGE military radar system was used
3	1960	The commercial airline reservation systems, semi-automatic business research environment went online with two connected mainframes
4	1962	J.C.R. Licklider was hired who introduced interest at ARPA and developed a working group called "intergalactic computer networks"
5	1969	A DARPA programs entitled "Resource Sharing Computer Networks" was initiated with certain objectives
6	1991	Home broadband was created
7	2001	To begin growing at a faster rate, home broadband enters into mainstream usage
8	2009	To show sequential and revenue growth, 10 GE was introduced
9	Today	100 GE standards are used

As the advantages of a star network is easy to wire, install and maintain. Another side, it requires more cable length and is more expensive than bus topology. The star networking is useful when some processing has to be centralized. *Bus network* (as shown in figure 2(d)) is used for temporary networks, easy extension and implementation. The drawbacks of bus network are that it is limited to a cable length and an easy fault in the cable can cause the destruction of the whole network [14]. This type of networks is mainly used for industrial applications. *Ring Network* is somewhat similar to bus network because it has no central host computer [15]. Each computer on this network has two neighboring nodes having their own applications independently [16]. It is in the form of a closed loop where each node can transmit the data by consuming the token as depicted in figure 2(e).

Broadcasting of data is quick in ring network, but as the data packets must pass through every computer between each source and destination because of that transmission, data is very slow and failure of any node can cause the unsuccessful data transmission. As all these wired networks are fixed and encounter with certain drawbacks such as they are non- portable, static in nature, drill the holes into the walls, handoff is minimum and require the cost of fiber+ copper +co-axial cable [17,18]. So to remove all such types of drawbacks, wireless networking came into existence. Moving of wired networks causes the whole rewiring and it is the biggest drawback of this technology. Peter Gold mark about 20 years ago introduced the concept of wired city i.e. the interconnection of telephones in the offices and between the offices, faxes, etc. A wireless networking is the one which uses high frequency radio signals instead of wires to communicate between nodes [19, 20]. The single hop and multi-hop are the two major types of wireless networks.

Single hop is a single connection between devices. Infrastructure less and infrastructure based are further extensions of single hop networking. *Infrastructure less* has no fixed structure between nodes as in Bluetooth. *Infrastructure based* has fixed structure like in cellular networks. *Multi-hop* is another type where two or more hops exist between each source and destination. Multi-hop is also categorized into infrastructure based and less infrastructure. Examples of infrastructure based are wireless sensor networks and wireless mesh networks and at least the example of infrastructure less is VANETS.

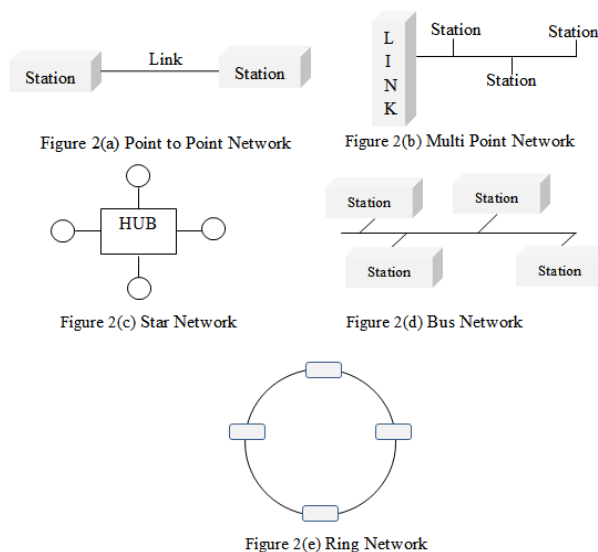


Figure 2. Point to Point and Multipoint Network

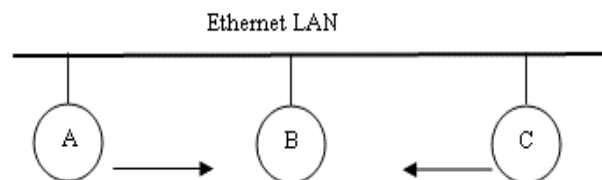


Figure 3 Ethernet network

Table 2 shows the difference between wired and wireless networks and the possible difference between Adhoc, WSN and WMN are depicted in table 3 also the pictorial differences of wired and wireless networks are shown in figure 3.

II. SECURITY

The cost of networking is continuing to decline and has become an essential part in completing *daily business tasks* [21]. Advancement in network technology has allowed for organizations to use the network not only to share

resources, but also to store large pool of data for analysis [22]. Therefore, securing such data and resources of organization on network is a great concern. As no computer network is completely secure.

TABLE II
DIFFERENCE BETWEEN WIRED AND WIRELESS NETWORKS

S.NO.	Wired Network	Wireless Network
1	Use network cable.	Uses radio frequency.
2	Allow faster and more secure communication.	Allow less secure communication.
3	Used up to 2000 feet distance.	The wireless network range is usually 150-300 indoors and up to 1000 feet outdoors
4	Traditional Ethernet communication offers only 10 Mbps bandwidth.	Wireless network supports a maximum bandwidth of 11 Mbps.
5	Inexpensive and Static in nature.	Expensive Mobility of wireless LAN helps offset the performance
6	Wired network connected to internet & firewalls is the primary security consideration.	Wireless network protects their data through wired equivalent privacy (WEP) encryption standard.

Security is generally defined as the state of being free from danger or threat [23]. The largest computer related crime in US history was committed by Kevin Mitnick which cost of 80 million dollars in US intellectual property [21].

TABLE III
DIFFERENCE BETWEEN AD-HOC, WSN AND WMN

Ad-hoc Network	Wireless Sensor Network	Wireless Mesh Network
Happens at OSI Layer 1	Nodes are stationary after deployment	It is a mesh topology with an assumption that every node in a network has a connection to every other node
All devices can directly communicate with any other devices within radio range	The aim is to prolong the lifetime of the network	The aim is to provide the services

The measurement of security for networks varies from situation to situation. The basic understanding about the security techniques is important for the research being done today. A web is subjected to attack from malicious sources and these attacks can be parted into two categories: passive attack and active attack [24, 25].

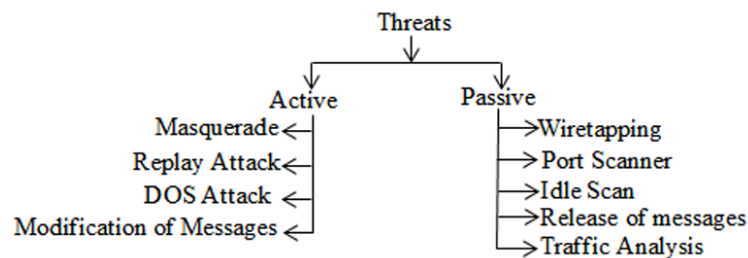


Figure 4 Possible Threats

Table 4 and figure 4 shows the explanation of possible threats and categories of active and passive attacks respectively. In *active attacks*, we have *masquerade* where one entity pretends to be a different entity. A *replay attack* occurs if the latter captures the message from the sender and receives the passive/replay message. In *modification of messages*, the altering and reordering is done by creating an unauthorized effect and at last in DOS, an attacker may suppress all messages sent to the receiver [28, 29]. If any organization's network is hacked, hackers may access all the individual databases of clients as easily as its employees [30]. Thus, the first thing to prevent your network secure is to grant the access only to authorized users. Figure 5(a) shows some intruders. An intruder may use several ways of gaining the access to your network. Some common intrusion methods are discussed in table 5.

TABLE IV
DESCRIPTION OF PASSIVE THREATS

S.NO.	Threat	Description
1	Wiretapping	The wiretapping is defined as the intruding of a telephone conversation by third party [26]. It is also known as telephone tapping, which is divided into two types- passive wiretapping monitors and records the traffic while active wiretapping is one which alters or affects the data.
2	Port Scanner	It is a software application which is designed specifically to probe a host for open ports [27].
3	Idle Scan	It is utilized to identify the availability of services by the attackers who send some spoofed packet to a computer.
4	Release of Message Content	In this procedure, the attacker's aim is to take the contents of the message transferred between sender and recipient.
5	Traffic Analysis	In this, the attacker's aim is to follow the pattern of message from source to destination.

To ward off such intrusions, organizations can consider some preventive methods (i.e. Firewall, encryption, antivirus software, etc.) as discussed in table 6. Security of network ensures the protection of whole networks. The entire field of security is too vast and is at evolutionary stage [32]. Security can be assorted into different categories [33] as described in figure 5 (b).

Some of are as Cyber security is security against cyber conjure and any crime with the involvement of computer and network is referring to as cybercrime. Data security is likewise known as infuse which is defined as defending the information from unauthorized access, disruption and alteration. Mobile Security is important in mobile computing and fixed as the security of personal and business information stored in smart phones.

TABLE V
DESCRIPTION OF PASSIVE THREATS

S.NO.	Intruder	Explanation
1	Trojan horse	A Trojan horse is a program similar to a virus which is applied to place the password information or just destroy the programs on the hard disk. Trojans often sneak in, attached to a free game.
2	Denial of Service	DOS are one of the worst attacks which is impossible to pass over. It is planned to bring the net to its knees by flooding the useless traffic
3	Email borne virus	These are those malicious codes which are transmitted as an attachment to your email. Thus, a source must be known before opening any attachment. To ward off an email borne virus, never lead a program unless sent by an empowered individual.
4	Packet sniffing	It is a program that captures the data (i.e. Username and password) for packets which travel over the networks. The highest exposure of packet sniffing is done at the cable modem because the entire neighborhood users are the constituent of any network [31].

Network security measures are required to protect the data during their transmission. Information Security means protecting the data from undesirable activities of unauthorized users. Different types of securities as described above exist in different layers of the network. eg. Network security occurs at the physical layer while data protection occurs at the application layer (as depicted in figure 6). During development of a strong network, the following security services are required to be considered i.e. access control, availability and non-renunciation. Figure 5(C) shows various security services. *Data confidentiality* is defined as keeping the privacy of the data [34]. *Integrity* means that the information transmitted by the sender is same as that of receiver [35].

Authentication means assuring the proof of identity [36]. *Non repudiation* means the power to demonstrate that the sender really sent the data or not [37]. *Access control* is the prevention of unauthorized usage of resources [38]. For any organization rather than to identify whether the network is wired or wireless, the main consideration is security assurance. Theoretically, wireless LANs less secure in comparison with wired LANs because signals travelling through the air in wireless LAN can be easily interpreted [39]. A secured wired network completes and complements a secure wireless mesh. To enforce the security of wired networks, certain patterns are required [40] e.g. Proper risk analysis, which deals understanding of risks relevancy makes a good wired security.

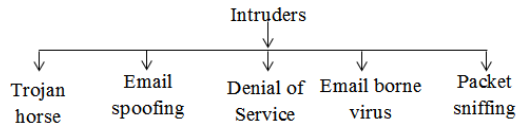


Figure 5(a) Intrusion Classification

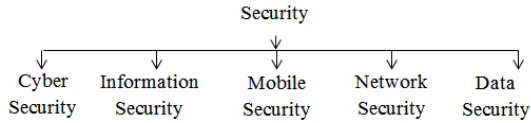


Figure 5(b) Types of Security

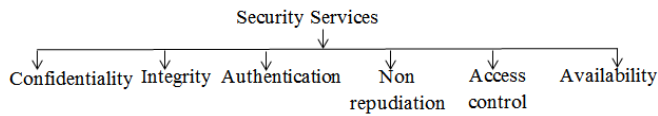


Figure 5 (c) security Services

Figure 5 Security Diagrams

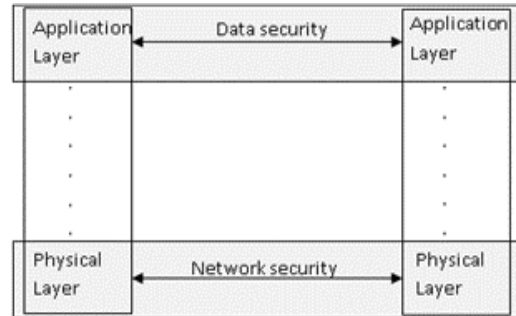


Figure 6 Data Security and Network Security

The Proper Network Security Policy must be placed to say that who may access the net, using what procedure network is accessed, what type of information is broadcast, and what encryption techniques are employed.

TABLE VI
PREVENTION EXPLANATION

S.NO.	Prevention Method	Explanation
1	Firewall	It is employed to supervise the information which is being transmitted to and fro to gain admission to a mesh
2	Encryption	Encryption is safer because even if a person is trying to intercept the information, he would not be able to understand it in any case. Encoding methods are more often than not applied to protect the confidentiality, integrity and authenticity of the mesh.
3	Antivirus software	An antivirus is a course of study which includes an auto update feature

Control the perimeter of your mesh is another important practice to know all the accesses points and to secure using proper firewalls, enforce the identification, authentication, confidentiality and integrity of the user and ultimately carry out a proper compliance monitoring mechanism [41].

TABLE VII
DIFFERENCE BETWEEN WIRED AND WIRELESS NETWORK

Wired Security	Wireless Security
Possible to secure the connection by hiding the wires inside the wall. End users have some confidence of validity of network they connect to. Less vulnerable in wired links.	Medium access is open to everybody within the range of transmission A user is completely unknown about the network connection More vulnerability in wireless links and Complicate trust relationship because of the flexibility of user mobility

Table 7 shows the difference between wired and wireless security. All these essentials discussed above are also supported for wireless networks, but with these, some other risks occur in wireless nets. In that respect are some additional links relevant to the wireless network as i) the installation of wireless network is inexpensive in contrast with wired networks. A wireless mesh can be set up easily without much technical knowledge because of this ease of induction as compared with a wired one, especially habitual risk analysis is not executed in a proper way. ii) In wireless networks, data is being sent through the air via radio waves. As the existence of wireless network is easier

to fix, an eavesdropper (who purely listen to the radio waves) can easily locate the wireless web. iii) The direct link between two workstations equipped with wireless network cards may have a dangerous risk for corporate networks. iv) Wireless Network can simply flood with static noise, then it is more comfortable to create a Denial of Service (DOS) attack and may cause the shutdown of a network [42, 43]. By creating a wireless network security police to secure a wireless net, it is a supreme starting point and ought to be gained by all the societies. Before operation of all access points, we must securely hold them. One should enforce the identification and authentication of a network as well as encryption and integrity of a mesh. The finding of the presence of any unauthorized access point is one of the most significant steps of ensuring protection. Noticing all the signals on a regular base, remote sensors must be employed to supervise the whole surroundings. To see an unauthorized access point, wireless security experts recommend 24x7 airwaves monitoring [44]. Such monitoring should identify unauthorized access point, unauthenticated traffic, off hour traffic and then along. During the development of wireless security, many attempts have been attained to achieve data confidentiality, integrity and common authentication. Initially used wireless security protocols are WEP, WPA and WPA2.

TABLE VIII
COMPARISON BETWEEN WPA, WPA2 AND WEP

S.NO.	Security	WPA	WPA2	WEP
1	Authentication	WPA pre-shared Key and WPA enterprise are used.	WPA personal and WPA enterprise is used.	WEP open and WEP shared key are used.
2	Data integrity	Message Integrity Code(MIC) using Michael algorithm	Cipher Block Chaining- Message Authentication Code(CBC-MAC)	Cyclic Redundancy Check (CRC)
3	Confidentiality	Temporal Key Integrity Protocol (TKIP)	Cipher block Chaining Message authentication code Protocol (CCMP)	Stream cipher mechanism RC4

WEP stands for Wireless Equivalent Protocol. The purpose of WEP is to provide security comparable to wired network. To furnish the data confidentiality, a common stream cipher mechanism, i.e. RC4 is used which encrypts the message with a shared key. In parliamentary law to provide data integrity, a common CRC (Cyclic Redundancy Checksum) is used which is IV4 (Integrity Check Value) and to provide authentication. WEP open system authentication and WEP shared key authentication is applied. WEP fails to fulfill the key management and it is a major drawback of it [45, 46]. **WPA** implements the majority of IEEE 802.11i standard and overcomes the flaws of WEP without requiring the new hardware. To provide data confidentiality, WPA adopts TKIP (Temporal Key Integrity Protocol) which still uses RC4, but also lets in a key mixing function. WPA introduces a weak keyed Message Integrity Code (MIC) using a Michael algorithm to improve the data integrity and at last authentication is provided by WPA pre-shared key (PSK) and WPA enterprise. The advantage of WPA is to offer key management through 4-way handshake and implements a sequence counter for replay protection [47].

WPA2 is advancement over WPA and implements the complete IEEE 802.11i standard. Data confidentiality in WPA2 is provided through a counter mode with Cipher block Chaining Message authentication code Protocol (CCMP) using AES (Advanced Encryption Standard). Cipher block chaining message code (CBC-MAC) is utilized to provide data integrity and at last authentication is provided using WPA2 personal and WPA2 enterprise. The major advantage of WPA2 is that it offers a robust key management using 4- way handshake and prevents from replay attack using 48-bit packet number [48]. Table 8 depicts the comparison between wireless secured protocol, i.e. WPA, WPA2 and WEP.

Today, network services (i.e. Email, www etc.) have become a basic need in day to day communication [48]. For providing these network services more effectively, WMN (Wireless Mesh Network) has sourced into a popular topology which builds high performance infrastructure. To supply a last mile broadband access, WMN is a promising technology. It is a most prominent movement of network propagation. Let us have a brief description on WMN, architecture, advantage and its applications.

III. HISTORY OF WIRELESS MESH NETWORKS

A. Wireless Mesh Network (WMN)

WMN is an extension of multi-hop Ad-hoc network and it is a combination of Ad-hoc and Mesh networking. Ad-hoc network is one where each device can directly communicate with any other device within its radio ranges while in mesh network each node acts as a router and has the capability to retransmit the packet to destination node [49, 50]. Figure 7 shows the scope of Wireless Mesh Network Technology.

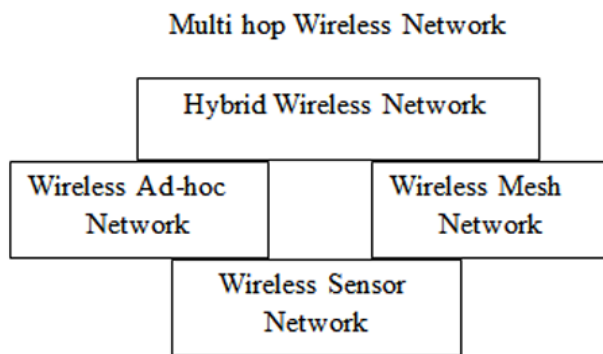


Figure 7 WMN Technology

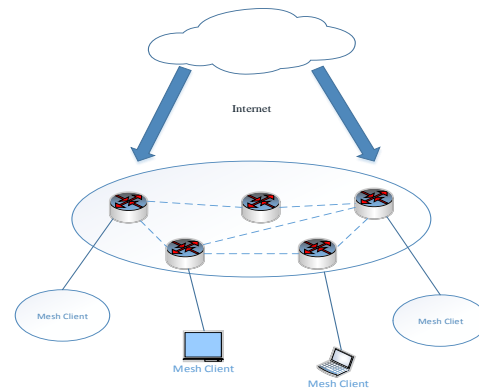


Figure 8 Infrastructure WMN

Similar to network classification, on the basis of connectivity, WMN is classified into three groups i) Point to Point, ii) Point to Multipoint and iii) Multipoint to Multipoint. Stop to Point networks are reliable, however their adaptability and scalability level is down. Point to Multipoint network has moderate scalability, but low reliability and adaptability. In order to surmount above limitations Multipoint to Multipoint networks are pioneered which endow with high reliability, scalability and adaptability [51]. The transmitting power of each node is scaled down as the number of clients in the mesh increases. To increase the coverage without need of transmitting power, Multipoint to Multipoint network uses multi-hop networking. The IEEE 802.11 family standards are used by Multipoint to Multipoint networks [52]. The networks which utilize these standards are called mesh networks and WMN are a particular class of Multipoint to Multipoint network. Table 9 shows the parametric difference between Point to Point, Point to Multi-point and Multipoint to Multipoint networks.

Multipliers can be especially confusing. Write “Magnetization (kA/m)” or “Magnetization (10^3 A/m).” Figure labels should be legible, about 10-point type.

B. WMN Architecture

Based on the functionality nodes, WMN architecture is classified into three main groups i) Infrastructure/backbone WMN ii) Client WMN and iii) Hybrid WMN. *Infrastructure/ backbone WMN formed by*

mesh routers for clients to relate them. Assorted cases of radio technologies are employed to make the backbone WMN. IEEE 802.11 is the most widely used technology, but in case of different radio technologies, clients must communicate with base station.

TABLE IX
WMN TYPES

WMN	Reliability	Adaptability	Scalability
Point to Point	High	Low	Low
Point to Multipoint	Low	Low	Moderate
Multipoint to Multipoint	High	High	High

Backbone WMN is the most commonly used WMN as all the networks of community or neighborhood can be built using infrastructure meshing [53]. In this, mesh routers are placed on the upper side to serve as the access point for users. The routers generally used two types of radios i.e. for backbone communication and for user communication. Figure 8 shows the Infrastructure WMN. *Client WMN* endows with peer to peer network amongst devices. In this, client nodes comprise routing as well as providing end user application to customers from a individual type of receiving set on devices [54]. Client WMN architecture is shown in figure 9. *Hybrid WMN* is a combination of above two i.e., backbone and client WMN. The network can be accessed by mesh client either through a network router or direct meshing with mesh client only. In this, the infrastructure of WMN provides connectivity to other network and clients routing capability provides improved connectivity and reporting [55].

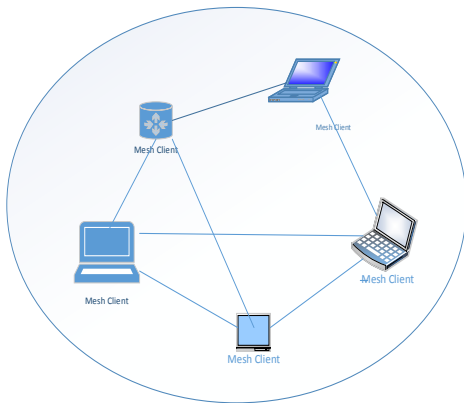


Figure 9 Client WMN

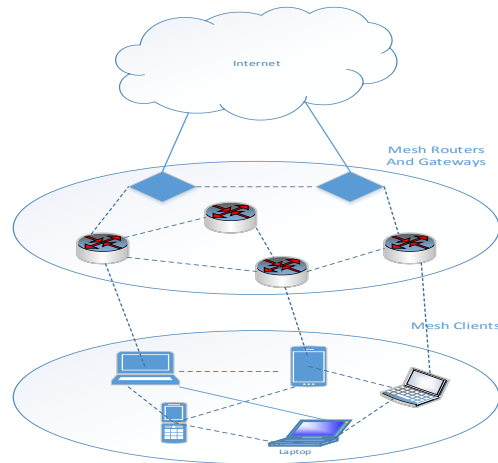


Figure 10 Hybrid WMN

Figure 10 shows hybrid architecture. Altogether the three architectures of WMN discussed above, consist of three types of lymph glands such as WMN client, WMN router and WMN gateway. *WMN client* is the end user devices that access the network for using the email, VoIP, gaming and location detection applications. The end user devices can be laptops, PDA's, smart phones, etc. The WMN clients have restricted power and routing capability [56, 57, 58]. It may or may not be connected to the network as its mobile nature. *WMN Router* is used to route the traffic of networks. The WMN mesh routers are reliable and possess a minimal consumption of transmission power. To enable the scalability in multi-hop mesh environment, multiple channels and multiple interfaces are utilized at the MAC in the chain of mesh routers [58]. *WMN Gateways* have the direct access to the internet. These are expensive in nature as they have multiple interfaces to connect to wired/wireless networks.

C. Benefits of WMN

WMN are less expensive than traditional networks and eradicate the installation cost of fibers and cables. For a larger coverage area, WMN is chiefly used [59]. WMN is Adaptable and Expendable and it can be added or taken away based on less or more coverage area. WMN is used where network configurations are blocked and how the lack of sight [60]. WMN **supports high demanding** indoor and outdoor connectivity and ideal to deliver high throughput and reliable connectivity. **Self-Configured and Self-Organized features** of WMN reduce the maintenance cost and setup time by enhancing the network performance [61].

D. Applications of WMN

Peer to Peer concept mesh topology helps to overcome the various deployment challenges such as installation of Ethernet cable, deployment models, etc. In case of path failure, the mesh topology concept results top quick reconfiguration of the path [62]. Mesh routers can be located anywhere as they are attached with freedom of mobility. These features of WMN draw the community to practice it in a diversity of applications. Some of them are given as below.

- **WMN in Smart Grids**

To enhance power savings or to upgrade the electrical infrastructure, smart power system is becoming a new global commercial enterprise. A smart power system is basically a rationalized electric grid, which offers authentic and effective distribution of electricity by using digital information and communication techniques [63]. It was brought out to minimize the costly environmental impacts and to ensure energy efficiency. Figure 11 shows the key concepts of smart grids.

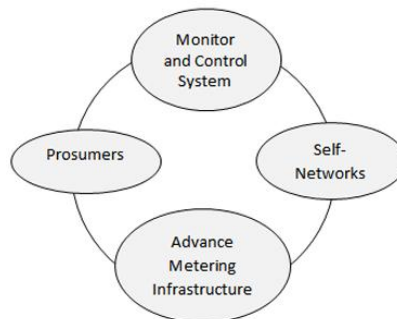


Figure 11 Smart Grid Concepts

- **WMN in Real Time Traffic Information Systems**

A probe technique is a method which collects real time traffic info. To transmit the traffic information data to the TMC (Traffic Management Centre), a feasible and cost efficient wireless communication is required. WMN is an architecture which is autonomous from any other wired/wireless network and needs low communication cost [64]. A WMN based traffic system consists of two components i) Probe vehicle during travelling on roads, probe vehicle automatically gathers real time traffic information and transmit to TMC over WMN. Probe vehicle is equipped with Data Collection Unit (DCU) and vehicular wireless terminals. ii) **WMN** consists of mesh clients and mesh routers. In this, mesh clients are our probe vehicles. WMN is formed dynamically by probe vehicle through wireless

connections. Information collected on vehicles reaches to the nearest mesh router and then mesh router communicates with TMC.

- *WMN in Motorola*

The mesh networking is a revolutionized wireless mobility. Mesh networking offers a seamless mobility in transforming wireless data communication for citizens and provides economic and safety benefits. Motorola has developed a MEA (Mesh Enabled Architecture) that enables cost effective, highly scalable network [65]. The carrying out of mesh networking is essentially performed in two modes i) infrastructure based- it is practiced to create wide or metro area networks. ii) Client meshes- it enables wireless networks. The singular feature of mesh architecture in Motorola is that links and routes are automatically formed between users. Motorola has launched various meshes enabled solutions that are Moto mesh (combines licensed and unlicensed radio in a single access point) and mesh track (provides accurate and fast user location).

- *WMN in Streaming Multimedia*

Multiple path existence between any pair of source and destination is one of the unparalleled qualities of WMN. A video file may have multiple replicas, if caching is included at nodes in WMN, as a result, if a new client request for a video file, it may get that file from multiple sources [66]. Whenever multiple clients are interested in various video files, then edifying a multiple multicast tree may not be the best choice. Instead of structuring a multiple multicast tree, existing multipath characteristic of WMN is more efficient. To establish a peer to peer streaming system and to find out the best video source location, let us presume that each WMN node has more or less memory space to save local copies and distribute these copies to peering WMN nodes. The connection status is periodically compiled by the media server. The server collects the file location information and preserves them in a DMT.

- *WMN in Cloud Computing*

Cloud computing is measured as an on-demand fifth utility application. The architecture of mobile cloud computing (MCC) is typically erect upon interest centric clouds, which allows the utilization of cloud services to mobile users. Traditionally MCC access suffers from high cost and WAN performance issues [67]. To reign over these issues, a mini cloud concept has emerged, known as cloudlets. A cloudlet is a local data center having the advantage of self-managing; faster access control, reduced cost in usage and deployment. By coalescing a cloudlet with a wireless connectivity i.e. WMN, local business can offer high performance cloud services to group MCC users. A WMN is a combination of two nodes, i.e. meshes router and mesh client which has the capability to establish mesh connectivity among them. Because of self- healing, adaptability and organizing features, WMN can espouse to topology during mobility and error recovery. Due to mobility management techniques, a mesh cloud architecture is being used which effectively supports transmission between network routers and gateways and potentially supports high bandwidth cloud services, low response time, reliability and so along. The integration of WMN and the mesh cloud framework offers self -organizing, self-management and flexible access to cloud services. As WMN is a fresh paradigm of wireless networking, it offers a fast, cheap and easy deployment of networks. Today, each organization is using this technology, so, it is the responsibility of WMN to provide services to users in a secured and effective manner [67]. One of the primary challenges of deploying these networks is a security matter. Altogether the above discussed applications as mentioned in the following systems.

IV. SECURITY ISSUES IN WIRELESS MESH NETWORK

A. Security Issues in Smart Grid

Single of the most requisite enabling components of the smart grid is the communication and network technology; merely in this there exist numerous scalability and protection matters. Security is one of the most acute anxieties in smart grids. It generally comes up during the pre-serration of confidentiality and integrity of smart metering data in AMI and other security consideration is a system monitoring and security of information.

B. Security Issues in Intelligent Transportation System (ITS)

The primary objective of ITS is to better the public safety by cutting down the accidents due to human mistakes. ITS technology has been steadily introduced in cars, but security is one of the major concerns in ITS. There exist two major security threats i) ITS security threat- It is the thread where hammers create bubbles around the vehicles to disrupt the receiving and transmission performance. ii) Wireless communication threat- DOS are the major type of threat, in this where network can be made unavailable by flooding the false messages that take up all the usable bandwidth. So cyber security should be done on availability, authentication and confidentiality.

C. Security Issues in Multimedia

The same security issues come up in multimedia i.e. confidentiality, integrity, authentication, availability, non-repudiation, accountability and encryption process are one of the major security threats.

D. Security Issues in Cloud Computing

As cloud computing offers an modern business for systems because of resilient, flexible, effective and scalability activities, governing bodies are still slow in admitting it. Several issues and challenges are allied with it. Security is one of the major challenges which harpers the growth of cloud. Security issues in cloud computing are i) data loss- a hacker might see your valuable data or might delete the target data. A data loss may occur when owner of data losses the key. ii) Account hijacked- if your account is hijacked by an aggressor, then it may employ the power of your reputation. An attacker having the control over account can eavesdrop the transaction, manipulate information, false damage response and hence along. iii) DOS- this attack is caught in rush-hour traffic, where customers will be placarded by the attacker's cloud service and there is no room to go to the destination except sit and await. Hence, the confidentiality, integrity, accessibility and accountability are major security attacks in cloud computing. Table 10 shows the security issues in different applications of WMN.

TABLE X
SECURITY CONCERNS IN VARIOUS APPLICATIONS

Applications	Security issues
Smart Grids	Authentication, Confidentiality, Integrity.
ITS	Availability, Authentication, Confidentiality
Multi Media	Confidentiality, Authentication, Integrity, Availability, Non-repudiation, Encryption process
Cloud Computing	Data loss, DOS, Integrity, Confidentiality, Accountability, Availability

V. SECURITY ISSUES AND TRENDS IN OSI MODEL

Open System Interconnection model (OSI Model) developed by ISO defines a networking framework to implement the protocols in seven layers. The OSI model helps to breakdown the networking function into seven layers [68]. The OSI seven layer model follows in order when computer leaves data while it conforms to its reverse

order when information enters into the data processor. The diagram of OSI model is shown in figure 12. The detailed explanation of all seven layers is discussed below.

A. *Physical Layer*

- *Responsibility*

It is responsible for frequency selection, carrier frequency generation, signal detection, modulation and data encoding. Existing wireless radios are able to support multiple transmission rates by combination of different modulation and coding rates. In society to increase the capability of wireless nets, various high speed physical technology have been invented i.e. UWB (Ultra Wide Band), OFDM [69]. To further increase the capacity, multiple antenna system has been used for wireless communication like antenna diversity and smart antenna technology, but due to high complexity and cost, fully adaptive smart antenna systems are used only in the base stations of cellular networks. In multiple antennas MIMO system are applied.

- *Attacks possible at Physical Layer*

The physical layer is responsible for signal detection, modulation and encryption of information. As Wireless Mesh Network (WMN) communicates through radio based medium, the most powerful attack at this layer is jamming attack [70]. Jamming attack impedes the radio frequencies. It is potent enough to dislocate the entire network communication [71,72]. If attacking devices do not obey MAC layer protocol then it is more intricate to detect them. The aim of jamming attack is to interfere in radio frequencies which are used during the communication in WMN. It may occur in three different ways i) jamming source: which disrupts the entire web. Ii) Less powerful jamming source: in this adversary potentially disrupts the network by passing around the jamming source. Iii) Intermittent jamming source: it proves unfavorable as some communication in WMN may be time sensitive.

- *Mechanisms Against various Attacks on Physical Layer in WMN*

The jamming attack can be fortified by employing different spread spectrum technologies:

In Frequency hopping spread spectrum, a pseudo random sequence is utilized which is known to both transmitter and recipient. By rapidly switching a carrier signal, signals are cast among many frequency channels. Thus, it is unmanageable for an attacker to predict the frequency selection, sequence and to jam it [73,74].

Direct sequence spread spectrum using a spreading code, each bit of original signal is characterized by multiple bits. Spread code spread the signal through a wide frequency band, to bring down the chances of meddling from any other tuner.

- *Physical layer open research issue*

The research issue includes improving the transmission rate and public presentation of physical layer technology through enhancement of the role of multiple antennas. MAC layer protocols required to be planned carefully to best use the advance feature provided by the physical layer.

B. *Data Link Layer*

- *Responsibility*

It secures the initial connection setup by dividing the information into data frames. DLL handles the recognition from a recipient that the data made it successfully. DLL divides the data packets into frames. Information packages

are encoded and decoded into bits [72]. It provides error handling, flow control and frame synchronization. The DLL is divided into two layers; i) MAC layer ii) LLC layer. The MAC layer control how a computer gains access to the data while the task of the LLC is to control the synchronization of frames.

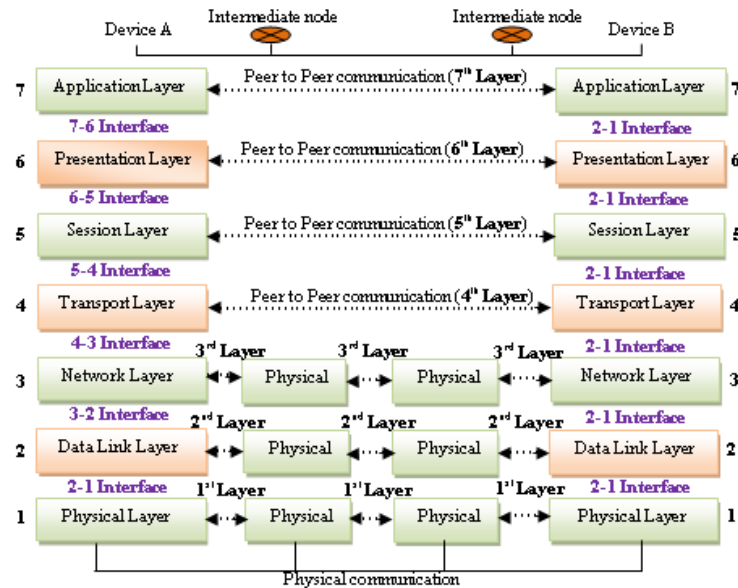


Figure 12 OSI Model Layer

• Attacks Possible at DLL

Jamming, eavesdropping, replays, MAC addressing and spoofing are some potential attacks on link layer of WMN. *Jamming attack* on link layer is more difficult to detect in comparison with the physical layer. In this, an attacker regularly transmits a MAC frame header on the channel so that, genuine nodes after the channel are busy may lead to denial of service attack [75]. In *Eavesdropping*, due to the broadcasting nature, wireless networks may prone to passive eavesdropping attack within the range of communication nodes. Passive eavesdropping does not immediately bear upon the functionality of the network, but conciliate data integrity and confidentiality. In Replay attack, another name of man-in-middle attack is replay attack. *Replay attack* can be launched by internal clients or external clients. If an attack is made by external nodes, then to reach the access over network resources, an attacker will transmit the messages at a later point of time, whereas an attack done by internal nodes, the attacker may keep copies of all data and to gain authorized access of resources [76].

• Security Mechanisms at Link Layer

To defend against frame collision attacks, various error-congestion codes were used and to provide the protection against passive eavesdropping data confidentiality service is used [77,78].

Based on permutation vector generation, Omari et al proposed a Synchronous Dynamic Encryption System (SDES). The SDES is robust against key compromise ii) integrity violation and biased bytes analysis. In this the security is ensured using two types of keys i) secret authentication key (SAK) ii) secret session key (SSK). Deng et al. Proposed a threshold and identity based key management, authentication scheme where key generation phase is

responsible for distributing the Pu/PR or master key for each client and authentication is realized by identity based mechanism. Another author proposed a wireless intrusion detection and response mechanism in which a system consists of a number of devices which are located near an access spot.

- *MAC layer research issues*

As scalability of WMN can be spoken by the MAC layer in 2 ways i) enhancing the existing MAC protocols or propose a new MAC protocol to increase close to end throughput. Ii) Allow transmission of multiple grooves in each mesh client for example CSMA/CA.

Thus, current open issues in the MAC are working on most of the existing MAC protocols based on CSMA/CA solve partial problems of overall issue, but raise other problems, i.e. how to essentially improve the scalability in multi-hop ad-hoc network.

C. Network Layer

- *Responsibility*

Routing and switching information is provided by the network layer. This layer makes a virtual circuit to transmit the data from node to node. The purpose of the mesh layer is addressing, internetworking, and error handling and congestion control.

- *Attacks at Network Layer*

Control packets and information packets are two cases of attacks on the network layer. These attempts are either dynamic or passive [79] in nature. Control packet attack targets the router functionality. The attacker's objective is to get to the route unavailable. Data packet attack targets the data forwarding functionality [80]. In this, an attackers aim is to basis the Denial of Service by injecting malicious data into the mesh. We first consider the control packet attacks, then mark data packet attacks.

Control packet attack, the first control packet attacks that targets on demand routing is rushing attack. In rushing attack, a route is requested from the root node to destination node by flooding the RREQ (Route REQuest) message with sequence numbers [80]. A delay is made between the receiving of the RREQ messages by a particular node and forwarding the nodes to next node. Attackers launch a malicious node between source and destination [81]. The intent of malicious node is to forward the RREQ message to target node before any other intermediate node. Thus, route between source and destination includes the malicious client, which then leaves out the packet of flow resulting DOS attack [82].

In Wormhole attack, the objective is same as rushing attack, but this can be accomplished by applying different schemes. In this more than one malicious node establish a tunnel between source and destination. So RREQ messages are forwarded between malicious nodes [83,84]. As between each origin and destination malicious nodes are included, it's up to the malicious node either to dismiss the entire parcel or some selective packets which are moving between source and goal.

In Black hole attack, As malicious node always replies for positive RREQ, then nearly all the dealings within a region of malicious node is aimed towards the malicious node [77, 78]. The result causes a DOS attack.

The Gray hole attack is a variation of black hole approach. The dropping of entire packets may lead to easy detection of malicious nodes. So, attacker introduced another attack, i.e. gray hole attack which may live undetected for longer duration of time by dropping selected packets [85, 86].

Data Packet attacks are primarily launched by selfish node. The most vulnerable attack in this is passive eavesdropping. In this the nodes are dependent on each other to forward the data [87, 88, 89, 90]. The selfish nodes may not perform data forwarding functionality. Selfish nodes either drop the selective packets or entire packets. The malicious node may introduce trash packets to increase the bandwidth or the packet process time of the network.

In Multicast Routing Attacks, an attacker's aim is to interrupt network communication by analyzing the traffic or leading to packet dropping [91, 92, 93].

- *Security mechanisms at Network Layer*

Authenticated Routing for Ad-hoc networks (ARAN) an on demand routing protocol is employed to provide an authenticated route discovery, setup, path maintenance. It supplies the security by using cryptographic certificates.

Process: The public key of the trusted certificate server is utilized where the key is known to all. Each node receives a certificate issued by the server whenever a node joins the network. The certificate carries the IP address of node, public key node, creation timestamp of certificate and expire time of the certificate. In this during the route discovery process, signed route discovery packet (RDP) is sent by a node which holds the IP address of the destination node, source node certificate, time stamp and a nonce. The node in the route discovery validates signature of previous node and removes the certificate of previous node after recording the IP address of it. The client signs the context of the packet, adds its own certificate signed by its individual key and transmits it to the forwarding node [94]. A route reply packet (REP) is created by destination node and unicast the packet along the same route. The REP includes the source IP address, certificate, nouns, timestamp, identifier of packet character. As REP reaches to the source node, it verifies the nuance and signature of the destination node. Whenever an attacker introduces a malicious, an error is generated because certificate of that node fails to establish the genuineness.

Drawbacks: If an attacker injecting a large bit of bogus control packets, then a node may not be able to verify the signature and force a node to discard some control packets.

Security-aware ad-hoc routing protocol (SAR) is unlike the traditional routing protocol, which exploits hop count, location metrics for setting the routing path, SAR uses trust values, relationships metrics among the nodes [95]. A client is able to process or forward the RREQ to next node only if it receives the required authorization or trust level. A shared secret mechanism or a key distribution mechanism is applied to determine the trust levels among the guests. Trust levels will not work at different security levels.

Drawbacks: To provide the security at different floors, a protocol needs different keys. As the number of keys increases at each level, its maintenance and stored computational overhead also increases.

Secure Routing Protocol (SRP) requires a security association (SA) existence between source and destination pair [96]. SA establishes a shared secret key between two nodes. Query sequence number (QSEQ) (used by destination to check validity of RREQ) and a random key identifier (QID) (to identify specific request) are transmitted by the source node. The source node's RREQ message is protected by MAC (Message Authentication Code) which is computed using shared key between source and goal. Each node forwards the received RREQ message, by adding

the identifier. The ranking of a query is maintained by all nodes. The rate generated queries have the highest precedence. At the destination node, after checking the validity of a query, destination node verifies the integrity and authenticity of a message and generates the RREP route replies using different paths. The integrity and authenticity of RREP are checked by the same process as RREQ.

Drawback: The modification of unauthorized routes by malicious clients cannot be prevented by SRP.

Secure Link State Routing Protocol (SLSP) operation is split into three parts; i) public key distribution and management (PKD) ii) Neighbor discovery iii) link state updates. PKD is used to transmit the public key certificates with zone while the NLP (Neighbor Lookup Protocol) is utilized to distribute the link state information [97]. The signed HELLO message (containing the sender MAC address and IP address) is used by NLP. The task of NLP is to generate a message notification to SLP about wary observations. Wary observations are those where a node claims the MAC address of the current node or the same MAC address is used by two different IP addresses. The initiating nodes' IP addresses are distinguished by link state updates (LSU). Whenever a client receives an LSU, it verifies its signature using a public key. The ranking priority of each neighborhood node is kept by each node; nodes with lower rates of LSU have the highest precedence. Whenever a malicious node floods spurious control packet in the mesh, due to generation of high rate traffic, the node will be attributed to lower priority and will never be included in the itinerary.

Drawback: it has higher computational overhead as there is a use of asymmetric key cryptography.

- *Network Layer Open Issues*

Routing protocols for WMN are different from those in wired network and the cellular net. Despite the accessibility of several routing protocols for adhoc networks, design of routing protocols for WMN is still an active research area for several reasons: network performance metrics need to be identified and used to better the operation of routing protocols. Scalability is the most critical question in WMN. Routing for multicast application. Cross layer design between routing and much protocol.

D. Transport Layer

- *Responsibility*

As data packets travel in the form of segments, the transport layer is responsible for end to end connectivity between source and goal. TCP and UDP are the two major protocols for transport layer [98]. Reliable data transport and real time delivery are two cases of protocols.

Reliable data transport is an ad-hoc transport protocols can be separated into two types: i) TCP variants ii) entirely new transport protocols. An enhanced version of TCP wired networks is TCP variants. TCP data and TCP ack take different paths in WMN which experiences different packet loss, latency and bandwidth. While in ATP transmission are rate based which achieves better performance [99]. Real time delivery is generally to provide end to end delivery TCP are used instead of UDP. Additional protocols, i.e. Real time protocol (RTP) and Real Time Transport Protocol (RTCP) are used for congestion control [100].

- *Possible attacks in transport layer*

SYN flooding attack, de-synchronization attack and session hijacking attacks are some potential attacks at the transfer layer. SYN flooding attacks are easy to launch at TCP. In this until resources required by each connection

are exhausted, an attacker may repeatedly make new connection request [101,102]. SYN Flooding Attack is a three way handshaking mechanism is applied to finish the session between two pairs of nodes as indicated in figure 13.

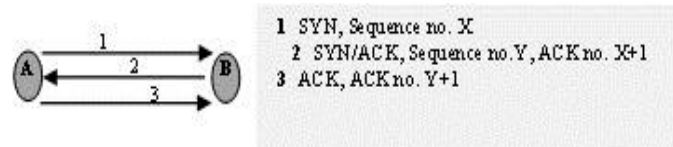


Figure 13 SYN Attack

Let assume if node A wants to establish a communication with node B then node A sends SYN packet along with a sequence number to node B. Node B sends SYN sequence, ACK sequence number finally node A completes the handshake process by sending ACK with ACK bit. Straight off, an attacker by sending too many SYN packets to node B may exploit and spoof the return SYN protocol. In *Security Hijacking*, the security mechanisms are offered only at established time, but not at the on-going session. Thus, an attacker may store the IP destination of a victim node and establishes expected, generated sequence number of victim node and then executes a DOS attack on victim node. Sequence number of victim node. In *de-synchronization attack*, the disruption of an existing connection refers to a de-synchronization attack. De-synchronization attack leads to a TCP ACK storm problem. In this, an attacker injects false messages by launching a session hijacking in an ongoing session between two clients. Ace of the communicating pair receives this false message and sends an ACK to another client. The other end node is not capable to distinguish the sequence number of this ACK so; it attempts to re-synchronize the session within its communicating peer. Thus, in this ACK packets go back and forth cause an ACK storm.

- *Security Mechanism at Transport Layer*

The protocols employed for securing transport layers are Secure Socket Layer (SSL), Transport Layer Security (TLS) and private communication transport (PCT) [103,104]. To secure the communication session, SSL/TLS use asymmetric key cryptography technique. EAP-TLS, an upper layer authentication protocol was proposed by Ababa and Simon. EAP-TLS offers mutual authentication between MR and MC. In this each terminal acts as an authenticator for its previous node.

- *Transport layer Research Issues*

It include several protocols exist for reliable data transport and real time delivery. Reliable data transport concerns with TCP data, ACK, ATP. In WMN, TCP data and acknowledgement follows different path results packet loss, bandwidth, and reaction time. Even if same path uses face network asymmetry problems, transmission in ATP is rate based. And for real time delivery protocols used are UDP, RTP and RCTP.

The current open research issues are working on is to avoid asymmetry between data and acknowledgement paths, it is hoped for a routing protocol to select an optimal route for both data and ack packets but without increasing the budget items.

E. Application Layer

Application layer supports the end user processes. It provides electronic mail; network software's and files transfer services [105]. Telnet and FTP are the applications that survive in this layer only.

- *Responsibility*

It ensures the user to access the network and provides the user support for services, i.e. email, network virtual terminal and file transfer [106].

- *Attacks at Application Layer*

Snooping attack and flooding attack are the two major attacks in WMN. Flooding attack affects the availability of victim as well as large portion of the network. While snooping attack affects the unity of the message being communicated [107].

- *Mechanisms*

Firewalls, IDS is most usual ways of securing application layer. Firewalls offer the protection against malware, spywares [108] etc. The brief explanation of OSI layer is read in the table 11.

TABLE XI
RESEARCH ISSUES AT EACH LAYER OF OSI MODEL

Physical Layer	Responsibility	Frequency selection, signal detection
	Attack	Jamming
	Mechanism	Frequency hopping spread spectrum, Direct sequence spread spectrum
	Research issue	New protocols need to be designed to use the advanced feature of the physical layer
DLL Layer	Responsibility	Framing the data packets
	Attack	Jamming
		Eavesdropping
		Replay
	Mechanisms	SDES
		SAK
		SSK
	Research Issue	Need to improve the scalability in multi hop ad hoc networks
Network Layer	Responsibility	Routing and switching information
	Attack	Control packet
		Data Packet
	Mechanism	ARAN
		SAR
		SRP
		SLSP
	Research issues	Routing for multicast applications
Transport layer	Responsibility	End to end data packet delivery
	Attack	SYN flooding attack
		Security hijacking
		De synchronization attack
	Mechanism	SSL
		TLS
		SSL/TLS
	Research issue	Avoid asymmetry between data and acknowledgement paths
Application layer	Responsibility	Ensures the user to access the network
	Attack	Snooping
		Flooding
	Mechanism	Firewalls
		IDS

VI. AUTHENTICATION PROTOCOLS

A. Responsibility

Authentication protocols are employed to assure the validity between mesh clients and mesh router before accessing the web servers.

B. Attacks on authentication protocols

Unauthorized access, in this an attacker may access the network resources by masquerading the legitimate client. *Spoofing attack* is used to form the MAC or IP address of legitimate node. In IP spoofing attack, attacker forwards

the packet by inserting the false source address or the destination of a legitimate node. While in *MAC spoofing*, the attacker modifies the transmitted MAC address originating from a legitimate node [109]. *DOS attack* is to create a buffer overflow by sending the flood of packets.

C. Security mechanisms against Attacks

Authentication mechanisms against these attempts are as given below: (a) Mishra et al suggested a standard mechanism for client authentication where users can access the mesh network without any device or software change [110]. (b) Cheikhrouhou et al proposed an architecture which is suited for multi hop WMN employing PANA. In this system, clients are authenticated on the production of cryptographic credentials [111]. (c) Prasad et al. given a light AAA infrastructure for continuous end to end security by deploying an AAA agent. The agent is a settlement agent for service providers [112]. Lee proposed a distributed authentication scheme for minimizing authentication delay by distributing the multiple trusted nodes over the network [113]. Above mentioned schemes are applied for wireless networks but given below are some authentication schemes which are used for WMN. (a) *ARSA* (Attack Resistant Security Architecture for Multi-hop WMN): ARSA [114] separates the whole network into multiple areas. Each domain is managed by a network operator or a broker. For accessing the network services, each MC has to register itself to its broker. After registration, broker will issue a universal pass to the clients. A network operator will allow the access service to only those customers which has valise universal pass. The significant factor for authentication in ARSA is universal pass. Fig 14 shows the diagram of ARSA.

AKES (an efficient authenticated key establishment scheme for WMN) like ARSA [115], mesh network in this is divided into number of domains. The goal of the scheme is to fix the session between mesh clients and between mesh client and network router. Fig 15 shows the diagram.

SLAB (A secure localized authentication and billing scheme for WMN [116] it ensures the security requirements by reducing inter domain handoff authentication latency and computation load.

LHAP (Light weight Hop by Hop access protocol): the protocol uses a light weight authentication [117] approach by authenticating the data packets. Intermediate nodes authenticate data packets before sending on to the adjacent hop.

Localized two factor authentication scheme for WMN Lin et al. Proposed [118] an inter-domain handover and mobility management in IEEE 802.11.

D. Authentication Handoff delay minimization

Due to the active nature of WMN, MC's may change their MR or domain from current access to new one. When an MC moves from one area to another, latency parameter needs to be centered. The movement of MC from current authenticated domain to new one requires re-authentication procedure which contributes to high delay in the network and encounters several types of approaches. A number of researchers have proposed various strategies in order to trim down the wait.

Li Xu and Yuan He. proposed a ticket based design to achieve a secure and fast handoff in WMN. Handoff security is provided by redistributing the tickets to MC and MR which authenticates each other without any third party involvement. The proposed design significantly reduces the communication and computational overhead latency.

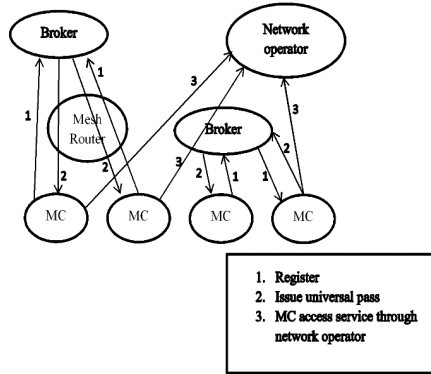


Figure 14 ARSA

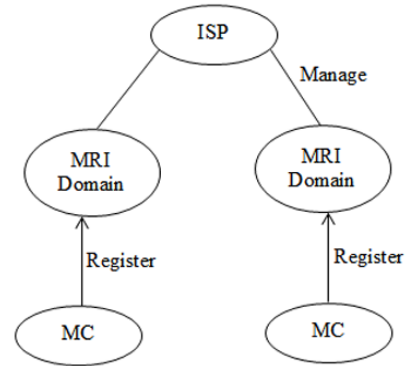


Figure 15 AKES

While T. Chen et al. Gives a security context transfer scheme [10] in which a context transfer activation request is transferred by previously access MR and the mobile user. Both MR and mobile user send a context activation request to the new MR. The authentication will be successful if the context activation request of both the parties is verified by the new MR. The depicted table 12 shows various handoff mechanisms with their intentions.

TABLE XII
Handoff Mechanisms

Approach	AIM/Purpose
Ticket based handoff authentication	Provide authentication in two phases
Security context transfer information	Cuts down handoff latency using security context activation request
Cluster chain based context mechanism	Reduces handoff in centralized WLAN
Fu et al.	Divide the mobile users into groups and proposes a group based handoff mechanism
Authors in [14]	Secure localized authentication and billing schemes for WMN
PKD	Proposes a proactive key distribution scheme using neighbor graphs
15,16	Uses a group key shared by all BS to support fast handoff

E. Current Authentication schemes

Security is a major challenge in creating the communication robust. From all the security services, authentication is a major concern due to the active nature of WMN. Any node may leave or join the network at any time. If a client enters into a network and wants to communicate with already existing node, firstly that a client needs to establish its identity. Various researchers have proposed different authentication systems in order to increase the security in WMN. Levente buttyan proposed a two certificate based system in which authentication process is carried out locally between MC and constraint MC. Computational operations are changed to MC in the former seat while short term certificates for digital signature are provided for later levels. The below table 13 indicates the current certification schemes.

TABLE XIII
AUTHENTICATION SCHEMES

Authentication Schemes	Purpose
Proxy encryption based secure multicast	Intermediate nodes transmit the messages without decrypting it
Centralized scheme	KDC generates a shared key and distributes it to the group
Decentralized scheme	The large community group is divided into small subgroups
SAKM	Dynamic and passive agents are utilized to furnish the protection
SIM-KM	The key control component is used for encrypting or decrypting the sender's and receiver's nodes
Fast certificate	Proposes two certificates based authentication systems

VII. OPEN RESEARCH CHALLENGES IN WMN SECURITY

WMN is considered as a new wireless network archetype, as it does not rely on any fixed infrastructure [119]. It is the focus of research in recent years owing to its great auspicious features. In order to access the services with stringent QOS requirements, recent researches of WMN focused on developing high performance routing protocols while the security and privacy issues of these protocols have acknowledged comparatively less attention. Due to dynamic and broadcasting nature of WMN, networks are vulnerable to a variety of attacks at almost each layer of the protocol stack. However, WMNs multi hop nature of communication may also lead to assorted cases of security attacks. Various researchers have proposed various security protocols, but there survive many challenges that necessitate to be spoken. Some of these challenges regarding security of WMN are discussed below:

WMN Handoff: The dynamic nature of WMN is one of the foremost reasons to come across it in recent researches. Mesh clients can move from one location to another and are able to access the network services after proving their legitimacy. The movement of a mesh client (i.e. roaming client) from its current serving mesh router (i.e. HMR) to the range of a new mesh router (i.e. Foreign Mesh Router FMR) is called handoff. The security issues that arise during handoff are:

Mobility Attacks: mesh clients are dynamic in nature, any client may leave its current serving mesh router and move to another router's range i.e. FMR. As the client moves from one location to another, an attacker may forge the IP address of a legitimate node and access the service by identifying itself as a legitimate node. So, the techniques which enhance the security of a roaming client are of the research topic.

Authentication Verification Delay: as a legitimate roaming client proves its identity to a FMR by following some authentication procedures, a significant delay in authentication verification process may enhance the security issues at a number of areas i.e. passive eavesdrop. It is the possibility of security getting the transmitted message between two legitimate nodes. It is always a great research issue in security. This attack may be compromised by enhancing the messages format communicated to the destination node. Even if the message is forged by a malicious node, it may not be able to read it anyway.

Centralized Authentication: Due to distributed architecture of WMN, it is difficult for a central authenticator to prove the authenticity of the entire nodes. The numbers of security issues at centralized authenticator are storage overhead (to store the security information or tickets of all the nodes), mobility management (to update the routing table to store the information of all the nodes).

Message Communication Cost: in a real time scenario, each communication message transmitted between entities is very expensive. A single communication message to prove the identity of a client takes a lot of cost. So, it is required to prove the validity of the node using minimum communication exchange.

Lack of Authentication Requirements: each authentication process needs to follow the necessary requirements regarding authentication mechanism i.e. i) authentication process should be fast enough to satisfy the QOS requirement of user services. ii) both MR and MC have to demonstrate their authenticity to each other and iii) protocols should be scalable (they must not degrade in performance as the mesh size increases). Existing authentication schemes for WMN undergo a slow authentication process which involves large latency that sometimes adversely affects the network services.

Attacks at Cross-Layer Framework: a cross layer framework may be regarded as a novel plan to improve the protocol efficiency. Generally, framework improves the network performance and the design can be attained in two ways: i) improving the performance at existing protocol layer or ii) merging the several protocols into one portion. In spite of being one of the current research topics, it guides to several drawbacks because the plan may be antagonistic against the existing protocols and may lose the abstraction of the protocol layer. For this, existing authentication techniques may merge to make a new protocol which is efficient against a variety of attacks. Use of novel techniques like merkley tree and graph based approaches; homomorphic encryption for secure multicast in MWN is a very interesting research trend.

The dynamic nature of WMN needs a protocol which dynamically changes the network topology based on links/nodes. Network coding is a technology to provide the solution over it. Although existing network coding protocols provide the security according to the network change of topology, but they may lead to the broadcasting nature of wireless medium which in turn is susceptible to a mixture of approaches such as eavesdropping and packet overflow attacks. So a future direction of this is to provide a protocol which is vulnerable against these attacks.

Based on the security issued discussed in this chapter, a bar chart is drawn which shows the percentage of work done at a particular field of security in figure 16. The depicted figure shows that maximum work is done over passive eavesdrop and handoff is the current ongoing research. Further communication cost and cross-layer framework are the two topics of future research.

VIII. CONCLUSION

This chapter gives an ample summary of networks from origin to today's technology i.e. WMN. After that an ephemeral discussion is abridged on WMN architecture, benefits, applications and research issues. The objective of this research is to deliberate the security mechanisms and issues that ascend in MWN. As entire communication is done through OSI model, this paper has discussed the research issues at each layer of OSI in WMN with their responsibilities, attacks and mechanisms. Finally the paper is summarized by giving the current research challenges in WMN security.

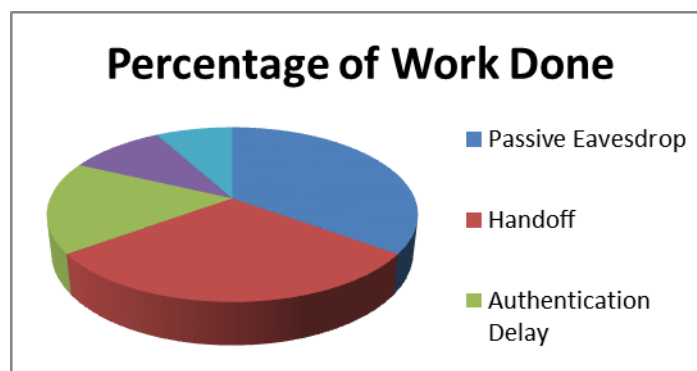


Figure 16 Work done Percentage

IX. FUTURE WORK

In future, peer to peer authentication delay can be considered as a significant research problem to be figure out. Peer to peer authentication means to affirm the identity of the source and destination node. In order to reduce

authentication delay, various advances have been suggested. A simulation over some authentication techniques has been made against authentication delay parameter a shown in table 14.

TABLE XIV
AUTHENTICATION DELAY APPROACHES

Number of Hops	Approach	Technique	Simulator	Simulation run	Average authentication delay
1-5 hops	AKES	Polynomial based key distribution	Qual-Net 4.5	10 times	100ms
1 hop	Fast Handoff	Ticket transfer	Qual-net4.5	10 times	150ms
1 hop	LAP	Data packet authentication	Qual-Net 4.5	10 times	237ms
1 hop	EAP-TLS		Qual-Net 4.5	10 times	249ms
1 hop	VLR-HLR	Authentication through third party	NS-2	10 times	800ms

In future research, authentication delay can be further reduced by employing one of the specific technique i.e. homomorphic operation using algebraic operations which enhances the security by applying computational operations on cipher texts.

REFERENCES

1. Norman Matloff, "overview of computer networks", 2005, retrieved on 7 Nov 2014, available at <http://heather.cs.ucdavis.edu/~matloff/Networks/Intro/NetIntro.pdf>.
2. http://en.wikipedia.org/wiki/Computer_network
3. katestanford, "a brief history of computer network technology", 2013, retrieved on 8 Nov 2014, available at " <http://visual.ly/brief-history-computer-network-technology>".
4. Joshua Muscatello, "a document on wired and wireless networks", 2005, reterived on 9nov,2014,availableat"<https://www.iup.edu/WorkArea/DownloadAsset.aspx?id=61283>".
5. Porter, Michael E., and Mark R. Kramer. "Creating shared value." *Harvard business review* 89.1/2 (2011): 62-77.
6. Alberts, David S., and Daniel S. Papp. *The information age: an anthology on its impact and consequences*. office of the assistant secretary of defense washington dc command and control research program (ccrp), 1997.
7. Arbaugh, William A., et al. "Your 80211 wireless network has no clothes." *Wireless Communications, IEEE* 9.6 (2002): 44-51.
8. Mentze, Duane, Nathan Harmon, and Brett Williams. "Wireless network communications methods, communications device operational methods, wireless networks, configuration devices, communications systems, and articles of manufacture." U.S. Patent No. 7,269,653. 11 Sep. 2007.
9. Ahlswede, Rudolf, et al. "Network information flow." *Information Theory, IEEE Transactions on* 46.4 (2000): 1204-1216.
10. Simpson, William. "The point-to-point protocol (PPP) for the transmission of multi-protocol datagrams over point-to-point links." (1992).
11. Highsmith, William, and John Wood. "Point-to-multipoint access network integrated with a backbone network." U.S. Patent Application 10/142,267.
12. Guimerà, Roger, et al. "Optimal network topologies for local search with congestion." *Physical Review Letters* 89.24 (2002): 248701.
13. Ebersole, Ronald J. "Ring bus hub for a star local area network." U.S. Patent No. 4,982,400. 1 Jan. 1991.
14. Ripeanu, Matei. "Peer-to-peer architecture case study: Gnutella network." *Peer-to-Peer Computing, 2001. Proceedings. First International Conference on*. IEEE, 2001.
15. Santra, Santanu, and Pinaki Pratim Acharjya. "A Study And Analysis on Computer Network Topology For Data Communication."
16. <http://www.csl.mtu.edu/cs4451/www/notes/Network%20Topologies.pdf>
17. Bla
18. Royer, Elizabeth M., and Chai-Keong Toh. "A review of current routing protocols for ad hoc mobile wireless networks." *Personal Communications, IEEE* 6.2 (1999): 46-55.
19. Balakrishnan, Hari, et al. "A comparison of mechanisms for improving TCP performance over wireless links." *ACM SIGCOMM Computer Communication Review*. Vol. 26. No. 4. ACM, 1996.
20. Johnsson, Andreas. *Bandwidth measurements in wired and wireless networks*. Department of Computer Science and Electronics, Mälardalen University, 2005.
21. Daya, Bhavya. "Network security: History, importance, and future." *University of Florida Department of Electrical and Computer Engineering* (2013).
22. <http://www.cs.iit.edu/~cs549/lectures/CNS-1.pdf>
23. Zeng, Ailin. "Discussion and research of computer network security." *Journal of Chemical and Pharmaceutical Research* 6.7 (2014): 780-783.
24. Kumar, Mahendra, and Ajay Bhushan Amit Kumar. "A Study of wireless Ad-Hoc Network attack and Routing Protocol attack." *International Journal of Advanced Research in Computer Science and Software Engineering ISSN* 2277 (2012).

25. William, Stallings, and William Stallings. *Cryptography and Network Security, 4/E*. Pearson Education India, 2006.
26. Jain, Kamal. "Security based on network topology against the wiretapping attack." *Wireless Communications, IEEE* 11.1 (2004): 68-71.
27. Kaushik, Atul Kant, Emmanuel S. Pilli, and R. C. Joshi. "Network forensic system for port scanning attack." *Advance Computing Conference (IACC), 2010 IEEE 2nd International*. IEEE, 2010.
28. Salem, Malek Ben, Shlomo HersHKop, and Salvatore J. Stolfo. "A survey of insider attack detection research." *Insider Attack and Cyber Security*. Springer US, 2008. 69-90.
29. Schuba, Christoph L., et al. "Analysis of a denial of service attack on TCP." *Security and Privacy, 1997. Proceedings., 1997 IEEE Symposium on*. IEEE, 1997.
30. Park, Kihong, and Heejo Lee. "On the effectiveness of probabilistic packet marking for IP traceback under denial of service attack." *INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*. Vol. 1. IEEE, 2001.
31. Ansari, Sabeel, S. G. Rajeev, and H. S. Chandrashekar. "Packet sniffing: a brief introduction." *Potentials, IEEE* 21.5 (2002): 17-19.
32. Danieli, Damon V. "Security services and policy enforcement for electronic data." U.S. Patent No. 6,510,513. 21 Jan. 2003.
33. Atkinson, Bob, et al. "Web services security (WS-Security)." *Specification, Microsoft Corporation* (2002).
34. Barrows, Randolph C., and Paul D. Clayton. "Privacy, confidentiality, and electronic medical records." *Journal of the American Medical Informatics Association* 3.2 (1996): 139-148.
35. Bishop, Matt. "What is computer security?." *Security & Privacy, IEEE* 1.1 (2003): 67-69.
36. Neuman, B. Clifford, and Theodore Ts'o. "Kerberos: An authentication service for computer networks." *Communications Magazine, IEEE* 32.9 (1994): 33-38.
37. Zhou, Jianying, and Dieter Gollmann. "Evidence and non-repudiation." *Journal of Network and Computer Applications* 20.3 (1997): 267-281.
38. Andress, Mandy. "Wireless LAN security." *Information systems security* 11.3 (2002): 29-33.
39. NETWORKSECURITY, TELECOMMUNICATIONSAND. "Wireless LAN Security."
40. Ashley, Paul, Heather Hinton, and Mark Vandenwauver. "Wired versus wireless security: The Internet, WAP and iMode for e-commerce." *Computer Security Applications Conference, 2001. ACSAC 2001. Proceedings 17th Annual*. IEEE, 2001.
41. Markham, Tom, and Charlie Payne. "Security at the network edge: A distributed firewall architecture." *DARPA Information Survivability Conference and Exposition*, Vol. 1. IEEE Computer Society, 2001.
42. Mohan, Devina. "Denial of Service attack in Wireless Mesh Networks."
43. Ben Salem, Naouel, and J-P. Hubaux. "Securing wireless mesh networks." *Wireless Communications, IEEE* 13.2 (2006): 50-55.
44. Perrig, Adrian, et al. "SPINS: Security protocols for sensor networks." *Wireless networks* 8.5 (2002): 521-534.
45. Lashkari, Arash Habibi, Mir Mohammad Seyed Danesh, and Behrang Samadi. "A survey on wireless security protocols (WEP, WPA and WPA2/802.11 i)." *Computer Science and Information Technology, 2009. ICCSIT 2009. 2nd IEEE International Conference on*. IEEE, 2009.
46. Bulbul, Halil Ibrahim, Ihsan Batmaz, and Mesut Ozel. "Wireless network security: comparison of wep (wired equivalent privacy) mechanism, wpa (wi-fi protected access) and rsn (robust security network) security protocols." *Proceedings of the 1st international conference on Forensic applications and techniques in telecommunications, information, and multimedia and workshop*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2008.
47. Arbaugh, William A. *Real 802.11 security: Wi-Fi protected access and 802.11 i*. Addison-Wesley Longman Publishing Co., Inc., 2003.
48. Kumkar, Vishal, et al. "Vulnerabilities of Wireless Security protocols (WEP and WPA2)." *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)* 1.2 (2012): 34-38.
49. Aggelou, George. *Wireless Mesh Networking*. McGraw-Hill Professional, 2008.
50. Srikrishna, Devabhaktuni, and Amalavoyal Chari. "Selection of routing paths based upon path quality of a wireless mesh network." U.S. Patent No. 6,965,575. 15 Nov. 2005.
51. Kodialam, Murali, and Thyaga Nandagopal. "Characterizing achievable rates in multi-hop wireless mesh networks with orthogonal channels." *IEEE/ACM Transactions on Networking (TON)* 13.4 (2005): 868-880.
52. Waharte, Sonia, et al. "Routing protocols in wireless mesh networks: challenges and design considerations." *Multimedia tools and Applications* 29.3 (2006): 285-303.
53. Akyildiz, Ian F., and Xudong Wang. "A survey on wireless mesh networks." *Communications Magazine, IEEE* 43.9 (2005): S23-S30.
54. Jun, Jangeun, and Mihail L. Sichitiu. "The nominal capacity of wireless mesh networks." *Wireless Communications, IEEE* 10.5 (2003): 8-14.
55. Bahr, Michael. "Update on the Hybrid Wireless Mesh Protocol of IEEE 802.11 s." *MASS*. 2007.
56. Campista, Miguel Elias M., et al. "Routing metrics and protocols for wireless mesh networks." *Network, IEEE* 22.1 (2008): 6-12.
57. Jun, Jangeun, and Mihail L. Sichitiu. "MRP: Wireless mesh networks routing protocol." *Computer Communications* 31.7 (2008): 1413-1435.
58. Sichitiu, Mihail L. "Wireless mesh networks: opportunities and challenges." *Wireless World Congress*. 2005.
59. Bruno, Raffaele, Marco Conti, and Enrico Gregori. "Mesh networks: commodity multihop ad hoc networks." *Communications Magazine, IEEE* 43.3 (2005): 123-131.
60. Jun, Jangeun, and Mihail L. Sichitiu. "The nominal capacity of wireless mesh networks." *Wireless Communications, IEEE* 10.5 (2003): 8-14.
61. Iannone, Luigi, et al. "Cross-layer routing in wireless mesh networks." *Wireless Communication Systems, 2004, 1st International Symposium on*. IEEE, 2004.

62. Raniwala, Ashish, and Tzi-cker Chiueh. "Architecture and algorithms for an IEEE 802.11-based multi-channel wireless mesh network." *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*. Vol. 3. IEEE, 2005.
63. da Silva Leitão, Ivo Tiago. "Wireless Mesh Networks for Smart-grids." (2012).
64. Zhang, Xuedan, et al. "A novel real-time traffic information system based on wireless mesh networks." *Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE*. IEEE, 2007.
65. Zhang, Yan, Jijun Luo, and Honglin Hu, eds. *Wireless mesh networking: architectures, protocols and standards*. CRC Press, 2006.
66. Andreopoulos, Yiannis, Nicholas Mastronarde, and Mihaela van der Schaar. "Cross-layer optimized video streaming over wireless multihop mesh networks." *Selected Areas in Communications, IEEE Journal on* 24.11 (2006): 2104-2115.
67. Li, Gang, et al. "A survey on wireless grids and clouds." *Grid and Cooperative Computing, 2009. GCC'09. Eighth International Conference on*. IEEE, 2009.
68. Handel, Theodore G., and Maxwell T. Sandford II. "Hiding data in the OSI network model." *Information Hiding*. Springer Berlin Heidelberg, 1996.
69. Nandiraju, Nagesh, et al. "Wireless mesh networks: current challenges and future directions of web-in-the-sky." *Wireless Communications, IEEE* 14.4 (2007): 79-89.
70. Xu, Wenyuan, et al. "The feasibility of launching and detecting jamming attacks in wireless networks." *Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing*. ACM, 2005.
71. E. Shi and A. Perrig, "Designing secure sensor networks", IEEE Wireless Communication Magazine, vol 11, no 6, pp. 38-43, 2004.
72. W. Xu, W. Trappe, Y. Zhang, and T. Wood, "The feasibility of launching and detecting jamming attacks in wireless networks", in Proceedings of the 6th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'05), Urbana-Champaign, IL, USA, pp. 46-47, May 2005, ACM Press.
73. Dong, Jing, Reza Curtmola, and Cristina Nita-Rotaru. "Secure network coding for wireless mesh networks: Threats, challenges, and directions." *Computer Communications* 32.17 (2009): 1790-1801.
74. W. Stallings, *Wireless Communication and Networks*, Person Education, 2nd edition, 2009.
75. Y. Law, M. Palaniswami, L. V. Hoesel, J. Doumen, P. Hartel, and P. Havinga, "Energy-efficient linklayer jamming attacks against wireless sensor network MAC protocols", in ACM Transactions on Sensor Networks (TOSN), vol 5, no 1, article no 6, February 2009.
76. J. Leu, R. Lai, H. Lin, and W. Shih, "Running cellular/PWLAN services: practical considerations for cellular/PWLAN architecture supporting interoperator roaming", IEEE Communications Magazine, vol 44, no 2, pp. 73-84, 2006.
77. A. D. Wood and J. A. Stankovic, "Denial of service in sensor networks", IEEE Computer, vol 35, no 10, pp. 54-62, 2002.
78. H. S. Soliman and M. Omari, "Application of synchronous dynamic encryption system in mobile wireless domains", in Proceedings of the 1st ACM International Workshop on Quality of Service and Security in Wireless and Mobile Networks (Q2SWinet'05), Montreal, Quebec, Canada, pp. 24-30, 2005, ACM Press.
79. A. Naveed, S. S. Kanhere, and S. K. Jha, "Attacks and security mechanisms", book chapter in: Security in Wireless Mesh Networks, Y. Zhang et al. (eds.), pp. 111-144, Auerbach Publications, CRC Press, USA, 2008.
80. Yuan, Yuan, et al. "ROMER: resilient opportunistic mesh routing for wireless mesh networks." *IEEE workshop on wireless mesh networks (WiMesh)*. Vol. 12. 2005.
81. Karlof, Chris, and David Wagner. "Secure routing in wireless sensor networks: Attacks and countermeasures." *Ad hoc networks* 1.2 (2003): 293-315.
82. Wood, Anthony, and John A. Stankovic. "Denial of service in sensor networks." *Computer* 35.10 (2002): 54-62.
83. Hu, Yih-Chun, Adrian Perrig, and David B. Johnson. "Wormhole attacks in wireless networks." *Selected Areas in Communications, IEEE Journal on* 24.2 (2006): 370-380.
84. Wang, Xia, and Johnny Wong. "An end-to-end detection of wormhole attack in wireless ad-hoc networks." *Computer Software and Applications Conference, 2007. COMPSAC 2007. 31st Annual International*. Vol. 1. IEEE, 2007.
85. Cai, Jiwen, et al. "An adaptive approach to detecting black and gray hole attacks in ad hoc network." *Advanced Information Networking and Applications (AINA), 2010 24th IEEE International Conference on*. IEEE, 2010.
86. Prathapani, Anoocha, Lakshmi Santhanam, and Dharma P. Agrawal. "Detection of blackhole attack in a Wireless Mesh Network using intelligent honeypot agents." *The Journal of Supercomputing* 64.3 (2013): 777-804.
87. Cai, Jiwen, et al. "An adaptive approach to detecting black and gray hole attacks in ad hoc network." *Advanced Information Networking and Applications (AINA), 2010 24th IEEE International Conference on*. IEEE, 2010.
88. Cai, Jiwen, et al. "An adaptive approach to detecting black and gray hole attacks in ad hoc network." *Advanced Information Networking and Applications (AINA), 2010 24th IEEE International Conference on*. IEEE, 2010.
89. S. Zhong, L. E. Li, Y. G. Liu, and Y. R. Yang, "On designing incentive-compatible routing and forwarding protocols in wireless ad-hoc networks: an integrated approach using game theoretical and cryptographic techniques", Wireless Networks, vol 13, no 6, pp. 799-816, December 2007, Kluwer Academic Publishers, Hingham, MA, USA.
90. N. Ben Salem, L. Buttyan, J.-P. Hubaux, M. Jacobson, "A charging and rewarding scheme for packet forwarding in multi-hop cellular networks", in Proceedings of the 4th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'03), Annapolis, MD, USA, pp. 13-24, June 2003, ACM Press.
91. A. Chen, D. Lee. G. Chandrasekaran, and P. Sinha, "HIMAC: high throughput MAC layer multicasting in wireless networks", in Proceedings of the 3rd IEEE International Conference on Mobile Adhoc and Sensor Systems (MASS'06), Vancouver, British Columbia, Canada, pp. 41-50, October 2006.

92. S. Roy, D. Koutsonikolas, S. Das, and C. Hu, "High-throughput multicast routing metrics in wireless mesh networks", in Proceedings of the 26th IEEE International Conference on Distributed Computing Systems (ICDCS'06), Lisbon, Portugal, p. 48, July 2006. IEEE Computer Society Press.
93. J. Sen, "Routing security issues in wireless sensor networks: attacks and defenses", book chapter in: Sustainable Wireless Sensor Networks, Y. K. Tan et al. (eds.), pp. 279-309, INTECH Publishers, Croatia, 2010.
94. K. Sanzgiri, B. Dahill, B.N. Levine, C. Shields, and E.M. Belding-Royer, "A secure routing protocol for ad hoc networks", in Proceedings of the 10th IEEE International Conference on Network Protocols (ICNP'02), Paris, France, pp. 78 – 87, November 2002.
95. S. Yi, P. Naldurg, and R. Kravets, "Security-aware ad hoc routing for wireless networks", in Proceedings of the ACM Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'01), pp. 299 – 302, Long Beach, CL, USA, October 2001.
96. P. Papadimitratos and Z. J. Haas, "Secure routing for mobile ad hoc networks", in Proceedings of the SCS Communication Networks and Distributed Systems Modelling and Simulation Conference (CNDS'02), San Antonio, TX, USA, pp. 27-31, January 2002.
97. P. Papadimitratos and Z. J. Hass "Secure link state routing for mobile ad hoc networks", in Proceedings of the Symposium on Applications and the Internet Workshops (SAINT'03 Workshops), pp. 379-383, Washington DC, USA, 2003.
98. Dierks, Tim. "The transport layer security (TLS) protocol version 1.2." (2008).
99. Karagiannis, Thomas, Andre Broido, and Michalis Faloutsos. "Transport layer identification of P2P traffic." Proceedings of the 4th ACM SIGCOMM conference on Internet measurement. ACM, 2004.
100. Iyer, Yogesh G., Shashidhar Gandham, and Subbarayan Venkatesan. "STCP: a generic transport layer protocol for wireless sensor networks." *Computer Communications and Networks, 2005. ICCCN 2005. Proceedings. 14th International Conference on.* IEEE, 2005.
101. B. Wu, J. Chen, J. Wu., and M. Cardei, "A survey on attacks and countermeasures in mobile ad hoc networks", book chapter in: Wireless Network Security, Y. Xiao et al. (eds.), pp. 103-135, Springer, Signals and Communications Technology, 2006.
102. A. D. Wood and J. A. Stankovic, "Denial of service in sensor networks", *IEEE Computer*, vol 35, no 10, pp. 54-62, 2002.
103. W. Stallings, *Cryptography and Network Security*, Fourth Edition, Pearson Education, 2010.
104. B. Aboba and D. Simon, "PPP EAP TLS authentication protocol", RFC 2716, 1999.
105. Banerjee, Suman, Bobby Bhattacharjee, and Christopher Kommareddy. *Scalable application layer multicast*. Vol. 32. No. 4. ACM, 2002.
106. T. M. Chen, G.-S. Kuo, Z.-P. Li, and G.-M. Zhu, "Intrusion detection in wireless mesh networks", book chapter in: *Security in Wireless Mesh Networks*, Y. Zhang et al. (eds.), pp. 146-169, Aurbach Publications, CRC Press, USA, 2008.
107. A. Mishra, K. Nadkarni, and A. Patcha, "Intrusion detection in wireless ad hoc networks", *IEEE Wireless Communications*, vol 11, pp. 48-60, 2004.
108. J. Sen, "An intrusion detection architecture for clustered wireless ad hoc networks", in Proceedings of the 2nd IEEE International Conference on Intelligence Communication Systems and Networks (CICSyN'10), pp. 202- 207, Liverpool, UK, July 2010.
109. Heberlein, L. Todd, and Matt Bishop. "Attack class: Address spoofing." *Proceedings of the 19th National Information Systems Security Conference*. 1996.
110. M. Kassab, A. Belghith, J.- M. Bonnin, and S. Sassi, "Fast pre-authentication based on proactive key distribution for 802.11 infrastructure networks", in Proceedings of the 1st ACM Workshop on Wireless Multimedia Networking and Performance Modeling (WMuNeP 2005), pp. 46–53, Montreal, Canada, 2005.
111. A. R. Prasad and H. Wang, "Roaming key based fast handover in WLANs", in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC 2003)*, vol 3, pp. 1570- 1576, New Orleans, Louisiana, USA, 2005.
112. O. Cheikhrouhou, M. Maknavicius, and H. Chaouchi, "Security architecture in a multi-hop mesh network", in *Proceedings of the 5th Conference on Security and Network Architectures (SAR 2006)*, Seignosse, France, June 2006.
113. M. Parthasarathy, Protocol for Carrying Authentication and Network Access (PANA) Threat Analysis and Security Requirements, RFC 4016, March 2005.
114. Y. Zhang and Y. Fang, "ARSA: an attack-resilient security architecture for multihop wireless mesh networks", *IEEE Journal of Selected Areas in Communication*, vol 24, no 10, pp. 1916–1928, 2006.
115. B. He, S. Joshi, D. P. Agrawal, and D. Sun, "An efficient authenticated key establishment scheme for wireless mesh networks", in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM'10)*, pp. 1-5, Miami, Florida, USA, 2010.
116. H. Zhu, X. Lin, R. Lu, P.-H. Ho, and X. Shen, "SLAB: A secure localized authentication and billing scheme for wireless mesh networks", *IEEE Transactions on Wireless Communications*, vol 7, no 10, pp. 3858–3868, October 2008.
117. L. Lamport, "Password authentication with insecure communication", *Communications of the ACM*, vol 24, no 11, pp. 770-772, 1981.
118. X. Lin, X. Ling, H. Zhu, P.-H. Ho, and X. S. Shen, "A novel localised authentication scheme in IEEE 802.11 based wireless mesh networks", *International Journal of Security and Networks*, vol 3, no 2, pp. 122-132, 2008.
119. Sen, Jaydip. "Security and privacy issues in wireless mesh networks: A survey." *Wireless networks and security*. Springer Berlin Heidelberg, 2013. 189-272

Performance-aware Cloaking Algorithm for Optimizing Anonymous Location-based Services

Dileep Kumar

Department of Information Media

The University of Suwon, Hwaseong-si South Korea.

Abstract—The prevailing infrastructure of ubiquitous computing paradigm on the one hand making significant development for integrating technology in the daily life but on the other hand raising concerns for privacy and confidentiality. As Location based services (LBS) equip users to query information specific to a location with respect to temporal and spatial factors thus LBS in general while Location Anonymizer, core component of privacy preservation models, in particular put under extreme criticism when it comes to location privacy, user confidentiality and quality of service. For example, a mobile or stationary user asking about his/her nearest hospital, hotel or picnic resort has to compromise their exact location information. Here in this paper we are addressing the significance of our proposed index optimized cloaking algorithm for Location Anonymizer with respect to performance, quality and accuracy which can be smoothly integrated into existing location anonymity model for privacy preservation. The main idea is to deploy R-tree based indexing scheme for Location Anonymizer to make best use of available computing resources. In accordance with the proposed approach, next step is to develop an index optimized cloaking algorithm which can cloak spatial region effectively and efficiently on behalf of R-tree based indexing scheme. Finally we will quantify the benefits of our approach using sampled results through experiments that the proposed cloaking algorithm is scalable, efficient and robust to support spatio-temporal queries for location privacy.

I. INTRODUCTION

Ubiquitous computing is the method of enhancing computer use by making many computers available throughout the physical environment, but making them effectively invisible to the user. LBS play a pivotal role in ubiquitous computing. Due to the proliferation of location based devices, prolific growth has been made to expand the huge base of LBS. For example, you might have observed the abrupt turning up of cabs at the stand since they have been using GPS devices, one of the most recent applications of LBS to track passengers and routes where the mobile user is willing to reveal his or her location. On the other hand, mobile users want to access LBS to locate nearest hospital without revealing his or her location as per privacy concern. Other examples include real time traffic congestion monitoring, detailed directions, integrated search results on dynamic maps, satellite imagery, location-based advertisements and etc. In location-based applications, location-based database server is responsible to process location-based query triggered by user with respect to the revealed spatial information [12]. In fact this authoritative location-based database server is untrusted primarily relying on the revealed user location, thereby raising critical concerns related to privacy and security of the registered users. In order to exploit location-based services, users have to compromise their privacy with such untrusted location-based database servers or simply ensure their privacy by limiting the quality of location-based services. Thus an adversary may access sensitive information related to user by breaching the security of untrusted location-based server or may reveal the user information by examining trends of different publicly available location-based data from such untrusted applications. In order to build up the confidence of users in location-based services, there is a pressing need to introduce such privacy preserving models that not only mitigate the privacy and security threats but also provide efficient and scalable computing mechanism. In this regard several cloaking algorithms have been proposed by research community to preserve user privacy for

motivating users towards location-based services [9]. In existing approaches [3, 2, 17, 20, 8], user is supposed to indicate his/her privacy requirement in terms of K-anonymity model which is then blurred into a spatial region by cloaking algorithm. Thus the efficiency of cloaking algorithm has direct impact on the characteristics of privacy preservation model with respect to quality, flexibility and privacy. Following are the shortcomings that motivated us for our contribution for privacy preservation towards location-based services. (i) Existing approaches employ cloaking algorithms on inefficient data structure such as hierarchical data structure [17] which seems to be a naive approach and supposed to exhibit poor performance in real time scenario when the scale and speed of mobile users are changed vigorously. (ii) Irrespective of being static or dynamic model [8, 16] several approaches lack efficiency in the absence of particular data indexing model. Our motivation to conduct this research study is to present a well balanced combination of cloaking algorithm and Location Anonymizer with underlying R-tree based indexing scheme which is efficient as well as privacy aware while the focus of existing approaches seems to skew towards privacy irrespective of overall efficiency of privacy preservation model. Our contribution made in this paper can be summarized as follows:

- Deploying R-tree based indexing scheme for location anonymizer to make best use of available resources i.e. memory and processing time
- Development of efficient cloaking algorithm to evaluate optimized result set in response to spatial queries of users
- Performance analysis of conducted experiments to demonstrate the effectiveness of our proposed algorithms

The structure of the paper is as follows. Section 2 presents related work in the area of LBS with respect to privacy preservation. Section 3 outlines the architecture of location anonymizer. Cloaking algorithm along with different practical scenarios is discussed in section 4. In section 5, we describe the performance evaluation of our proposed cloaking algorithm in contrast with other existing approaches. Finally, we conclude our work in section 6.

II. RELATED WORK

The K-anonymity model [18] is one of the premier models in the data privacy domain which have influenced research community unanimously. In the most general sense, K-anonymity model addresses the concern of privacy when it comes to release a version of private data by data holder for practical usage with some scientific assurance that privacy of individuals cannot be compromised by matching similar trends of different version of data. L. Sweeney [20] defines K-anonymity as: A relation is said to be K-anonymous provided each record in the relation is indistinguishable from at least K-1 other records with respect to a set of quasi-identifier attributes. In the context of Location Based services, K-anonymity model ensures the privacy of mobile users by making their locations indistinguishable among at least other K-1 users. The synergy of K-anonymity model with spatio-temporal cloaking [9] is one of the open debates among researchers interested in emerging privacy preservation models. Thus K-anonymity is one of the crucial requirements for flexible and efficient location-based query processing model [3] as widely discussed in Clique-Cloak algorithm [2], spatio-temporal cloaking algorithm [9] and peer-to-peer spatial cloaking algorithm [5]. The Clique-Cloak algorithm [2] can only accommodate few users with limited K-anonymity requirements due to computation overhead and also suffer from topographical adversary attacks. The spatio-

temporal cloaking [9] technique fails when it comes to scalability since this technique is optimized to track each single movement individually. The peer-to-peer spatial cloaking algorithm [5] finds K-1 NN(Nearest Neighbor) of query source but this approach suffers from "center-of-ASR" attack. Mokbel et al [6] presents algorithm for privacy preservation in scenarios where user can reveal its location for public queries while can also hide both location and query for private query at private location. The addressed algorithm works on snapshot queries as well as continues queries. The significant part of this work is the identification of centralized location anonymizer for location-based services implying a hierarchical data structure which is supposed to be invulnerable to adversary attacks such as query sampling for snapshot queries and query tracking for continues queries. This scheme is subjected to poor performance in the presence of hierarchical data structure in real time scenario and also proposed cloaking algorithm fails to defend maximum movement boundary attack and public queries versus private queries attack. Prive [8], a distributed system for query anonymization in LBS, comes with HilbASR algorithm. Prive only considers snapshot queries, lack of support for continuous queries and also static version of HilbASR lacks flexibility in terms of data indexing model. Our approach is quite similar as that of Casper, query processing paradigm discussed in Mokbel et al [16]. Casper employs hierarchical pyramid based data structure at Location anonymizer which is quite a naive approach with respect to performance and scalability of overall system. Thus the performance of cloaking algorithm over such a naive approach is not much impressive. Apart from that it requires user to continuously report anonymizer with the current user location thus requires frequent updates to maintain the overall pyramid based data structure for mobile users. In some mobile user distribution scenarios, Caspers hierarchical partitioning method fails to provide anonymity. Last but not the least efficiency of Casper query processing model can be improved by deploying some indexing scheme for location anonymizer. Our proposed scheme for privacy preservation is also centered across K-anonymity model to provide efficient, flexible and scalable cloaking algorithm for location-based services and it distinguishes itself from all previous schemes in performance as well as in scalability by implying indexing scheme at location anonymizer to ensure robust, accurate and fast query processing.

III. SYSTEM ARCHITECTURE

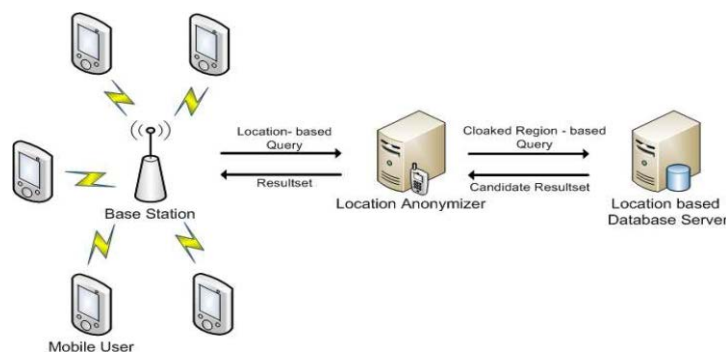


Figure 1. System Architecture

In general, our proposed scheme can be deployed using system architecture shown in Figure 1. The system architecture has three main components: the mobile user, the location anonymizer and the location-based database server. The mobile user interacts with the system by registering privacy profile [7] which specifies the typical privacy requirement of user with respect to the K-anonymity (K) and the minimum area (Δ). K-anonymous

parameter of privacy profile simply indicates that among how much other users the particular mobile user do not want to be distinguishable, while $_$ specifies desired coverage of the cloaked spatial region. Larger values of and results in restrictive privacy thereby degrading quality of service. Different privacy profiles can be set by mobile users to meet the desired privacy level at any time. Next step for mobile users is to report their spatial location and/or spatial query to the location anonymizer. The location anonymizer, after receiving location updates from mobile users, employs spatial cloaking technique to blur the users location and/or queries into cloaked spatial region as per specification of privacy profile. The blurred cloaked spatial region is then sent to the location based database server which is spatially tuned to deal with the spatial cloaked region instead of exact point location, at this back-end of system the candidate result set rather than actual result set is computed according to cloaked spatial region. The candidate result set is then returned to the mobile user through location anonymizer which previously computed the blurred spatial region based on the privacy profile, eventually mobile user who initiated the query is responsible to extract actual result from the candidate result set. The efficiency and accuracy of proposed system directly influenced by data structure scheme and cloaking algorithm embedded in the core of system i.e. the location anonymizer and the strictness of privacy profile. In the next section, we will discuss how our proposed location anonymizer can really play its part to drive the overall system efficiently and accurately as far as underlying data indexing and cloaking algorithm is concerned. On the other hand, the strictness of privacy profile simply puts operational trade-off between the level of privacy and the QoS (Quality of Service) which is solely manifested by the mobile user.

IV. LOCATION ANONYMIZER

The location anonymizer is a trusted third party that acts as the middleware between mobile users and back-end location based database server. The location anonymizer incrementally keeps track the number of users residing in the system and also consistently keep track of continuous movement of mobile users. Therefore, a key question for developing location anonymizer is: How accurate, scalable and efficient is the cloaking mechanism employed by location anonymizer? In order to address this primary concern for location anonymizer, we propose for the deployment of an indexing scheme to address mobile users efficiently and effectively.

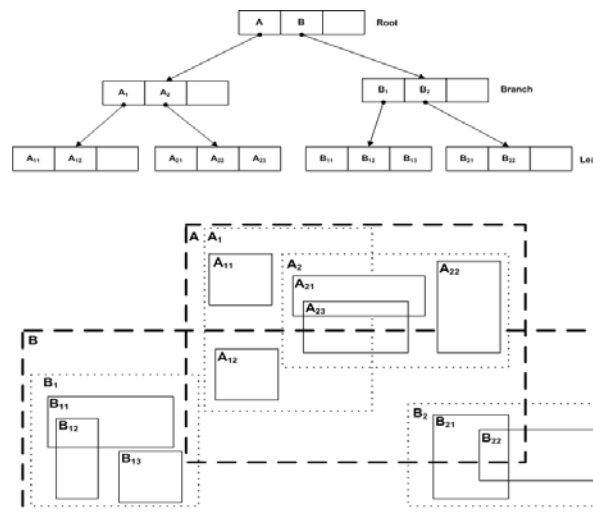


Figure 2. R-tree based Location Anonymizer

The fundamental assumption used for designing efficient location anonymizer is that the spatial datasets are indexed by the structures of R-tree family. R-tree and its variants [11] are considered excellent choices for indexing a range of spatial data such as points, line segments, rectangles, polygons and so forth and have already been deployed commercially (Informix and Oracle). The R-tree [10] is one of the most popular approaches for spatial access methods i.e., for indexing multi-dimensional information such as (x,y) coordinates of geographical data. R-tree is hierarchical in nature where high level node is termed as MBR (minimum bounding rectangle) that supposed to enclose a set of child MBRs at the next level in hierarchy except the lowest level where data objects are stored within the MBRs as depicted in the Figure 2. Root node represents the coverage of whole space of system and then the overall system space is hierarchically decomposed. Until now a number of R-tree variants [15] have been developed by the research community [19, 4, 13, 14, 1] as for as optimized performance of this promising indexing scheme for real-world data is concerned. R+ trees [19], a dynamic index for spatial access methods, differ from conventional R-tree by avoiding overlapping of internal nodes by inserting an object into multiple leaves thereby improving performance of point query since all spatial regions are covered by at most one node as well as fewer path traversal. Beckmann et al. [4] proposed the R*-tree which is more efficient in insertion and space utilization than the R-tree based on force-reinserted technique to avoid the overhead of splitting a full node. The Hilbert R-tree [13], which uses a Hilbert space filling curve outperforms the R*-tree by giving a better space localization. SR-tree "Sphere/Rectangle-tree" [14], an index structure for high-dimensional nearest neighbor queries, differs from other R-tree variants by combining utilization of bounding spheres and bounding rectangles thus showing improved performance on nearest neighbor queries by reducing both the volume and the diameter of regions. Conventional R-tree does not promise good worst-case performance but performing well when it comes to real-world data. The Priority R-Tree [1] is one the efficient R-tree variants and is at the same time worst-case optimal. Irrespective of different R-tree variants, in our proposed model the whole space can be decomposed into several spatial regions. The root of R-tree will represent the whole projected space. Each entry within a leaf node stores two pieces of information related to an element i.e. (i) Bounding box identifier of that element (ii) Number of users in corresponding bounding box. When a mobile user is registered in the system with some unique identifier and privacy profile, a hash table is maintained to store an entry of form (uid,prf,bid) where uid is the user id, prf is user defined privacy profile and bid is the bounding box identifier for bounding box holding that particular user. Any of existing R-tree scheme can be employed to search, insert and delete elements in the data structure.

One of the crucial aspect of location-based application is to maintain the mobile user updates as such applications tend to operate in highly dynamic environment thus requires flexible data structure for frequent updates. When a mobile user changes its location, location update is sent to location anonymizer in the form (uid,x,y) where uid is the user-identifier, x and y are new spatial coordinates after location update. In order to get new bounding box for the updated location of user, location anonymizer simply applies hash function $h(x,y)$. Now depending on the resulting bounding box, if it matches with the previous bounding box then location anonymizer will not do any processing at all. In case if it does not match with the previous bounding box then location anonymizer performs three operations i.e. updating new bounding box, incrementing user in new bounding box and decrementing user in old bounding box. If a new user is registered, a new entry will be marked in hash table and then marked in R-tree by insertion

procedure. If a user quits the system then its corresponding entry in hash table is deleted and then removed from R-tree by deletion procedure.

V. CLOAKING ALGORITHM

Algorithm 1 depicts the cloaking mechanism implied by location anonymizer. Input to the cloaking algorithm is the user privacy profile prf and bounding box identifier bid . Here we assume that bounding box is large enough to accommodate privacy requirement

Algorithm 1 Cloaking Algorithm

```

1: Function Cloaking ( $prf, bid$ )
2: if  $bid.N \geq prf.K$  and  $bid.Area \geq prf.Amin$  then
3:   return  $spRegion(bid)$ ;
4: end if
5:  $s \leftarrow 1$ 
6:  $bids = Sibling(bid)$ ;
7: while  $bids \neq Null$  do
8:    $N = N + bid.N + bids.N$ ;
9:    $A = bid.Area + bids.Area + A$ ;
10:  if  $N \geq prf.K$  and  $A \geq prf.Amin$  then
11:    return  $spRegion(bid, s)$ ;
12:  else
13:     $bids = Sibling(bid)$ ;
14:     $s \leftarrow s + 1$ 
15:  end if
16: end while

```

V. CONCLUSION

The horizon of LBS is being diminished by the raising concerns for privacy and confidentiality. This paper presents an index optimized cloaking algorithm for location anonymizer which can be deployed not only to motivate mobile users towards LBS without compromising privacy but also boost the overall performance of the spatial system for location privacy with respect to quality and efficiency. Existing cloaking algorithms for privacy preservation model in LBS [16] is suffered by immature data structure scheme employed for Location Anonymizer thereby exposing poor performance through deployed cloaking algorithm. Apart from that performance of these existing schemes can further be improved by implying indexing technique such as R-tree. Our proposed scheme can be deployed in centralized as well as distributed environment and is free from all existing adversary attacks by devising robust, efficient and flexible cloaking algorithm for privacy preservation.

ACKNOWLEDGMENT

This research supported by “THE UNIVERSITY OF SUWON, SOUTH KOREA”.

REFERENCES

- [1] L. Arge, M. de Berg, H. J. Haverkort, and K. Yi. The priority r-tree: a practically efficient and worst-case optimal r-tree. In SIGMOD

- '04: Proceedings of the 2004 ACM SIGMOD international conference on Management of data, pages 347–358, New York, NY, USA, 2004. ACM.
- [2] L. L. B. Gedik. A customizable k-anonymity model for protecting location privacy. ICDCS, 2003.
- [3] R. Bayardo and R. Agrawal. Data privacy through optimal k-anonymization. In ICDE, pages 217–228, Washington, DC, USA, 2005. IEEE Computer Society.
- [4] N. Beckmann, H.-P. Kriegel, R. Schneider, and B. Seeger. The r*-tree: An efficient and robust access method for points and rectangles. In SIGMOD Conference, pages 322–331, 1990.
- [5] C.-Y. Chow, M. F. Mokbel, and X. Liu. A peer-to-peer spatial cloaking algorithm for anonymous location-based service. In GIS '06: Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems, pages 171–178, New York, NY, USA, 2006. ACM.
- [6] C. C.Y. and M. Mokbel. Enabling private continuous queries for revealed user locations. In SSTO, volume 4605 of Lecture Notes in Computer Science, pages 258–275. Springer, 2007.
- [7] S. Duri, J. Elliott, M. Gruteser, X. Liu, P. Moskowitz, R. Perez, M. Singh, and J.-M. Tang. Data protection and data sharing in telematics. Mob. Netw. Appl., 9(6):693–701, 2004.
- [8] P. Ghinita, G. Kalnis, and P. Skiadopoulos. Prive: Anonymous location-based queries in distributed mobile systems. WWW, 2007.
- [9] M. Gruteser and D. Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. MobiSys, 2003.
- [10] A. Guttman. R-trees: a dynamic index structure for spatial searching. In SIGMOD '84: Proceedings of the 1984 ACM SIGMOD international conference on Management of data, pages 47–57, New York, NY, USA, 1984. ACM.
- [11] N. M. Hee Kap Ahn and H. M. Wong. A survey on multidimensional access methods. Technical report, 2001.
- [12] C. S. Jensen, A. Friis-Christensen, T. B. Pedersen, D. Pfoser, S. Saltenis, and N. Tryfona. Location-based services: A database perspective. In ScanGIS, pages 59–68, 2001.
- [13] I. Kamel and C. Faloutsos. Hilbert r-tree: An improved r-tree using fractals. In VLDB '94: Proceedings of the 20th International Conference on Very Large Data Bases, pages 500–509, San Francisco, CA, USA, 1994. Morgan Kaufmann Publishers Inc.
- [14] N. Katayama and S. Satoh. The sr-tree: an index structure for high-dimensional nearest neighbor queries. SIGMOD Rec., 26(2):369–380, 1997.
- [15] Y. Manolopoulos, A. Nanopoulos, A. N. Papadopoulos, and Y. Theodoridis. R-trees have grown everywhere. Technical report, 2003.
- [16] M. Mokbel, C. Chow, and W. Aref. The new casper: Query processing for location services without compromising privacy. VLDB, pages 763–774, 2006.
- [17] M. Mokbel, C. Chow, and W. Aref. The new casper: A privacy-aware location-based database server (demonstration). ICDE, 2007.
- [18] P. Samarati. Protecting respondents' identities in microdata release. IEEE Transactions on Knowledge and Data Engineering, 13(6):1010–1027, 2001.
- [19] T. K. Sellis, N. Roussopoulos, and C. Faloutsos. The r+-tree: A dynamic index for multi-dimensional objects. In VLDB '87: Proceedings of the 13th International Conference on Very Large Data Bases, pages 507–518, San Francisco, CA, USA, 1987. Morgan Kaufmann Publishers Inc.
- [20] L. Sweeney. Achieving k-anonymity privacy protection using generalization and suppression. Int. J. Uncertain. Fuzziness Knowl.-Based Syst., 10(5):571–588, 2002.

Encrypting grayscale images using S_8 S-boxes chosen by logistic map

Tariq Shah and Ayesha Qureshi*

Department of Mathematics, Quaid-i-Azam University, Islamabad, Pakistan

Abstract

In the present manuscript, we will design an encryption algorithm for grayscale images that is based on S_8 S-boxes transformations constructed by the action of symmetric group S_8 on AES S-box. Each pixel value of the plain image is transformed into $GF(2^8)$ with a dissimilar S_8 S-box chosen by using the logistic map. In this way, there are 40,320 possible choices to transform a single pixel of the plain image. By applying the generalized majority logic criterion, we will establish that the encryption characteristics of this approach are superior to the encoding performed by AES S-box or a single S_8 S-box.

Keywords: AES S-box, S_8 S-boxes, logistic map, generalized majority logic criterion.

1 Introduction

In the recent world, when more and more sensitive data is stored on computers and transferred over the internet, we require to guarantee information security and refuge. The image is also an important constituent of our information. Hence, it is really important to protect our image from illegal access. There are so many algorithms on hand to guard image from illegal access [1].

In [2], Joan Daemen and Vincent Rijmen developed the Rijndael block cipher which was adopted as Advanced Encryption standard (AES) by the National Institute of Standards and Technology (NIST). It was published as FIPS 197 in 2001 [3]. Being the only nonlinear transformation S-box is a pivotal component of AES. The security of the algorithm strongly relies on the strength of the S-box. Therefore, many researchers have paid their attentions to the improvement of S-box. In [4], 40,320 new S_8 S-boxes are obtained. These S-boxes are constructed by applying the permutations of symmetric group S_8 on the elements of AES S-box. All the good cryptographic characteristics of AES S-box are inherited in S_8 S-boxes.

In this work, we will design an encryption algorithm for grayscale images using S_8 S-boxes transformations selected by using the logistic map. Various parameters derived from statistical analysis used by the generalized majority logic criterion are calculated for the encrypted image and the strong points of proposed approach of using multiple S_8 S-boxes as compared to AES S-box or a single S_8 S-box are observed. They include correlation, entropy, contrast, homogeneity, energy, and mean of absolute deviation (MAD). Also, we will justify the effectiveness of S_8 S-boxes in proposed image encryption algorithm by using the generalized majority logic criterion.

This paper is organized as follows: In Section 2, we present a general summary about S_8 S-boxes and logistic map. In Section 3, we give details of the proposed image encryption algorithm. Section 4 consists of analysis of majority logic criterion for S_8 S-boxes and its comparison with AES S-box and a single S_8 S-box. Finally, we conclude in section 5.

2 S_8 S-boxes and the logistic map

2.1 S_8 S-boxes:

In [4], Hussain et al. has constructed S_8 S-boxes by using the action of symmetric group S_8 on Advanced Encryption Standard (AES) S-box [1]. In this way 40,320 S-boxes are received. The elements of $GF(2^8)$ are input of an S-box and the output always takes a unique value in $GF(2^8)$, in this way the S_8 S-boxes are bijective. The other cryptographic characteristics of AES S-box are also inherited in S_8 S-boxes.

2.2 Logistic map:

Replacing the logistic differential equation, $\frac{dx}{dt} = rx(1-x)$, with the quadratic recurrence equation, $x_{n+1} = rx_n(1-x_n)$, where $r \in (0,4)$ and x_0 is the initial value in the interval $(0,1)$, gives the so called logistic map. This quadratic map is capable of very complicated behavior. It is frequently cited as a model case of how complex, chaotic (disordered) behavior can arise from very simple nonlinear dynamical equations. This map is well known, one dimensional discrete map which reveals chaotic behavior for several values of the parameter r . In general, the logistic map exhibit chaotic behavior for the values of r on the interval $(3.56995,4)$ [5].

3 The proposed image encryption algorithm

The proposed image encryption algorithm is a two-step procedure. The first step includes the propagation of a sequence of indexes of S_8 S-boxes using the logistic map. The second step transforms each pixel of the plain image using a dissimilar S_8 S-box according to the generated sequence of indexes to encrypt the image.

3.1 The image encryption procedure

The image encryption algorithm is keyed out as below:

- (i) Transform the plain grayscale image I into a matrix.
- (ii) Convert each pixel of I into binary.
- (iii) Generate a finite sequence $a_i, i = 1, 2, 3, \dots, 40320$, of indexes of S_8 S-boxes by using the logistic map $x_{n+1} = rx_n(1-x_n)$ as: $a_i = \text{rem}(\text{round}(x_{n+1} \times 10,000,000), 40320) + 1$, with the parameter values of $r = 3.9$ and $x_0 = 0.75$.
- (iv) Apply S_{a_i} -box transformation on the pixel at position a_i , the ordering of the pixels is by row. The left most four bits of the pixel are used as a row value and the right most four bits are utilized as a column value. These row and column values serve as indexes into the S_8 S-box to select a unique 8-bit output value.
- (v) Transform the modified matrix into image to hold the encrypted image I' . Fig. 1 explains the operation of encryption procedure.

3.2 The image decryption procedure

The decryption algorithm is same as that of encryption algorithm with the replacement of step (iv), that is, apply S_{a_i} -box transformation on the pixel at position a_i , the ordering of the

pixels is by row. Find the pixel value in the S_{a_i} -box, the corresponding row and column values serve as the 8-bit output value.

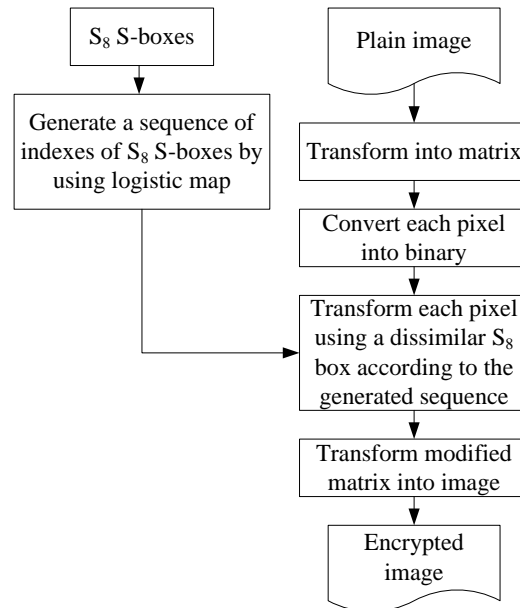


Fig. 1 Operation of encryption procedure

4 Experimental results and analysis

In this segment, we consider a grayscale picture of size 512×512 pixels to assess the functioning of our proposed algorithm. The visual analysis of Fig. 2 divulges that the proposed algorithm is more expert at hiding the data contained in it than that of using AES S-box or a single S_8 S-box. In summation, it indicates the possibility of using the algorithm successfully in both encryption and decryption.

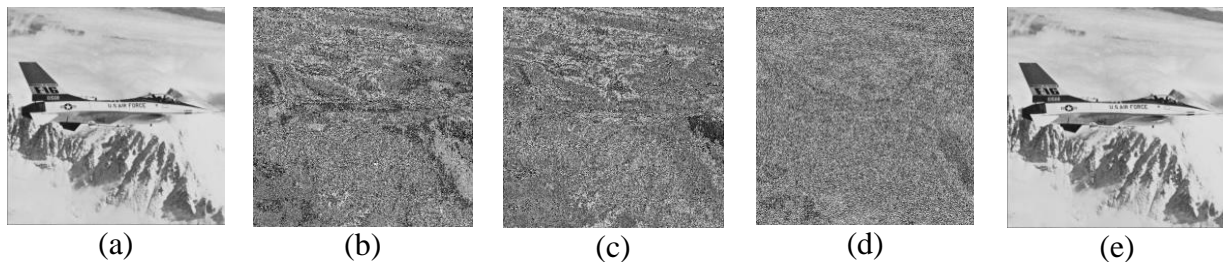


Fig. 2 (a) The plain image, (b) Encrypted image using AES S-box, (c) Encrypted image using S_8 S-box, (d) Encrypted image using S_8 S-boxes (proposed algorithm), (e) Decrypted image using proposed algorithm

The histograms in Fig. 3 shows that the encrypted image (d) bears no statistical resemblance to the plain image (a), and hence does not provide any clue to employ any statistical attack on the proposed image encryption procedure. Moreover, the histogram (d) is more homogeneous than that of (b) and (c), this fact adds strength to the proposed algorithm.

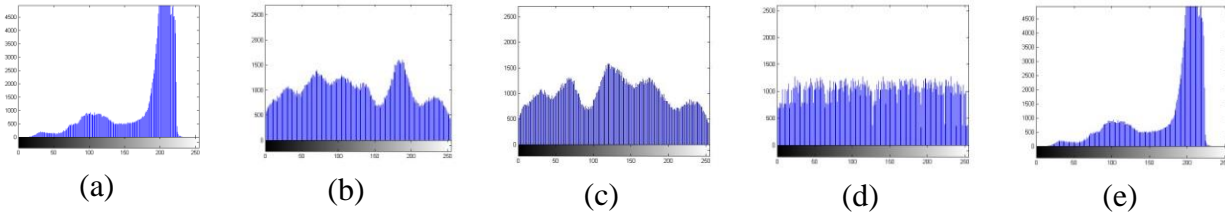


Fig. 3 Histograms: (a) The plain image, (b) Encrypted image using AES S-box, (c) Encrypted image using S_8 S-box, (d) Encrypted image using S_8 S-boxes (proposed algorithm) (e) Decrypted image using proposed algorithm

In table 1, we have derived several statistics used by the majority logic criterion that provide information about the texture of the encrypted image. Table 1 indicates that the amount of correlation, homogeneity and energy of the encrypted image (d) by using the S_8 S-boxes according to the proposed algorithm is smaller than that of the encrypted image (b) by using the AES S-box and the encrypted image (c) by using a single S_8 S-box. Whereas: the values for contrast, entropy and MAD are greater for the encrypted image (d). Hence, by majority logic criterion we can conclude that the presented approach of using S_8 S-boxes in image encryption is more secure.

Table 1 Analysis

Attribute	Fig. 1		
	AES S-box [6]	S_8 S-box [6]	Proposed algorithm
Contrast	6.0793	5.9696	9.8742
Correlation	0.2228	0.2721	0.0200
Energy	0.1397	0.1333	0.0157
Homogeneity	0.5832	0.5828	0.3966
Entropy	6.7975	6.8053	7.9714
MAD	65.54359	66.8351	65.8935

5 Conclusions

In this study, a grayscale image encryption algorithm based on S_8 S-boxes transformations constructed by the action of symmetric group S_8 on AES S-box is presented. The conduct of this algorithm is similar to the substitution box like encryption algorithms. Each pixel value of the plain image is transformed into $GF(2^8)$ with a dissimilar S_8 box chosen by using the logistic map. The suggested algorithm is tested for its encryption strength by the role of statistical analysis used by the majority logic criterion. The outcomes indicate that the performance of the suggested algorithm is superior to the encryption performed by AES S-box or a single S_8 -box.

References

- [1] M. Khan and T. Shah, A literature review on image encryption techniques, 3D Research 5(4), 1-25 (2014)
- [2] J. Daemen and V. Rijmen, AES proposal: Rijndael, AES Algorithm Submission (1999)
- [3] J. Daemen and V. Rijmen, The design of RIJNDAEL: AES-The Advanced Encryption Standard, Springer-Verlag, Berlin (2002).

- [4] I. Hussain, T. Shah and H. Mehmood, A new algorithm to construct secure keys for AES, International Journal of Contemporary Mathematical Sciences 5(26) 1263-1270 (2010)
- [5] R. M. May, Simple mathematical models with very complicated dynamics, Nature 261, 459-467 (1976)
- [6] I. Hussain, T. Shah, M. A. Gondal and H. Mahmood, Generalized majority logic criterion to analyze the statistical strength of S-boxes, Z. Naturforsch 67a, 282-288 (2012)

A Dynamic Media Access Control Protocol For Controlling Congestion In Wireless Sensor Network By Using Fuzzy Logic System And Learning Automata

Foroogh Karimi and Mansour Esmaeilpour*

Department of Computer Engineering, Hamedan Branch, Islamic Azad University, Hamedan, Iran

ABSTRACT

One of the existing layers in the reference model whose designing is of particular complication is the control layer of access to MAC media, it's proper designing causes to reduce interference and consequently to reduce energy consuming and to increase the network efficiency. In the recommended method, our focus is on the networks being multi-channel in order to distribute the network current through the different channels. In the first step of the research, we have used a layering structure for a better management of the network so that we could prevent congestion via the network management. This management is performed through using Fuzzy logic system logic system. The output of our Fuzzy logic system is the election of the best and most appropriate choice in order to continue route finding. But if a congestion of one incident takes place, we possess learning automata for assigning the channel searchingly for balancing the channel current. Using the resemblance maker of NS2, the results of the resemblance-making maintain that the recommended method has improved more greatly than the two basic protocols and could achieve the quality parameters of route finding services.

Keyword

Wireless sensor networks, Congestion control, Multichannel, Fuzzy logic system, Learning Automata

1. INTRODUCTION

Wireless sensor networks are a type of networks that are organized automatically using a certain number of wireless sensors. Using the different methods, they collect the received data from the environment based on which they analyze the environment. Wireless sensor network applications include supervising the environment, smart spaces, smart areas, medical areas, robotic discover, agricultural section, and ultimately military areas [1]. The industry also has too much tendency to use wireless sensor networks. Many researchers are seeking for the better ways to promote the characteristics and endurance of these wireless sensor networks [2]. But an important layer existing in stack protocol whose appropriate designing has caused to reduce interference and thereby to diminish improved energy-consuming and to increase the network efficiency is called the access control layer to the media. In some previous introduced plans, the protocol worked well with a few nodes in some networks, but by increasing the number of nodes, the network effectiveness was decreased [3]. Therefore, for such situation present tries not to let the protocol effectiveness reduced in the crowded networks. The main purpose of using several channels for MAC protocol includes minimizing the internal interference, to avoid the external interference, and to improve the network throughput. [4, 5]

2. RELATED WORK

The prevalent MAC protocols [6] in the wireless sensor networks have generally restricted themselves to use one channel such QMAC [7]. Such protocols usually work well in the networks which are of low traffic and are not in the exposure of an external noise emanated from the adjacent network frequencies. In this study, the researcher has made her best to prevent the reduction of the protocol effectiveness in the crowded networks. Therefore, we have benefited from the protocols that use several channels [8, 9].

2.1 TMCP

TMCP [10] is a multi-channel, tree-based protocol that has been designed in order for data gathering applications from the wireless sensor network. The main idea of TMCP protocol is categorizing the network into some separate sub-trees all of which are branches from the sink node so that the different channel are assigned to each sub-tree. Then each data current is led to the sink just through its own sub-tree. Competition and breakdown in the tree branches are

problems that have not yet been solved. The problem of this procedure is that the more we approach to the sink, the more time amount the existing nodes spend in this area to achieve the time slice channel and the more latency they undergo. Assembling also cannot be undertaken in TMCP because the relationships among the nodes in the different sub-trees are closed. Since TMCP assigns the fixed frequencies to the nodes, it causes some problems such as the “absent receiver” or the lack of listener. MAC algorithms enjoy multi-channel advantages such as MMSN in [11], HyMAC in [12] and YMAC in [13] all of which have been presented in the article but they do not use energy more effectively.

2.2 Queen-MAC Protocol

Time slice is used in the protocol [14] in such a way that the nodes wake up in this time slice and evaluate the transition media to exchange the data. The sensor node are in asleep state when the time slice is insufficient reducing the number of alertness slice, the recommended protocol induces a saving in energy. In this protocol, the alertness continuity for each sensor node can be determined considering its traffic load. Dygrid system is used to determine the slices called the record number slice into which the sensor nodes must wakeup. In this method, layering is used to control nodes. The recommended protocol uses some channel so that it could send some frames simultaneously. Assigning channels is performed so simply in this protocol and it possesses overhead.

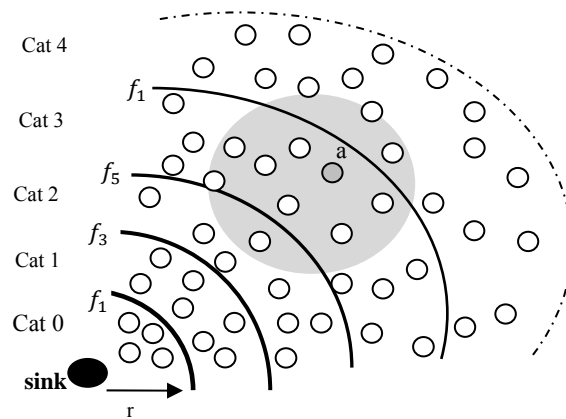
3. THE PROPOSED PROTOCOL

The network hypotheses:

- The network consists of many sensor nodes.
- The sink node is accessible to all next nodes.
- The network nodes do not have the movement capability.
- The normal nodes of the network excluding sink lack the geographical site-sending system.
- The amount of the normal nodes energy in network is limited.

Considering the methods of routing and because of the speed and dynamics of exploratory routing method, we have chosen this method to conduct a research study. Subsequently, in order to implement the decisions a hypothetical layering method was used. In order to use the decision-making process to deliver routing packages to the sink in the first phase and to assign the channel

in the second phase of the project, a structural distinction must be made. Using concentric circles can divide the network into different groups of nodes, so that the necessary decisions can be made.



(Fig 1: A view of the layering system)

To do the layering, first the sink node sends a signal to all nodes and each node that receives the signal react to this signal at a time with respect to the distance it has with the sink. The nodes are placed in appropriate layers according to the distance they have with the sink. After layering the nodes, a node such as 'a' which is displayed in the figure by a different color, intends to send its data packet. In the proposed method, at first, each node lists its single-phase neighbors, then, by sending periodic (scheduled) is informed of the status of the available nodes. The prerequisite of sending the packet in the first place, is sending it to the lower layer. However, if in the single-phase process, the node in the lower layer was not available, the node would transfer the packet to a node within the same layer, which is the most qualified node to do the transferring task. The ratio of the qualification of a node, in this case, is defined as a node which is in the Δt time period and has minimal possible traffic load, the shortest distance, the shortest queue length and also has the highest energy level. This selection is done using the proposed Fuzzy logic system. The inputs of the Fuzzy logic system are energy, traffic load, distance and queue length of the node, and the output is the value of every node. Therefore, at this stage, the interference and increase of the load on specific nodes in the network is reduced and the death of the first network node will be postponed.

The area of the zero categories is equal to $1\pi r^2/4$ and for category 1 is equal to: $G_1 = 1/4\pi(2r)^2 - G_0$ and through the same procedure the area for the other categories are calculated. And in general, the area of the G_i categories is equal to $G_i = (2i + 1)/4 \cdot \pi r^2$. The rate of the G_1 area to G_0 will be equal to $G_1/G_0 = 3$ and likewise the rate of the ' $i+1$ ' area to the ' i ' area will be $G_{i+1}/G_i = 2i + 3/2i + 1$. This means that a sensor node in the category i , by average, is responsible to transfer the traffic of $2i + 3/2i + 1$ nodes from the $i+1$ category. For example, if each node needs to send x data unit to send its report, then for instance a node in the category 3 according to figure 4-1, in spite of transferring its x data unit, is also responsible to transfer $9x/7$ data units of category 4 nodes, which is totally $16x/7$. And in the same way, a node in category 2 will send its x data units as well as $16x/7$ of data units of category 3. Therefore, with the assumption that each sensor node in the network uses x units of data in average for sending a report, it can generally be said that in the network that is composed of ' n ' categories of sensor nodes, a node in the ' i ' category has the responsibility to transmit and direct F_i data units according to the following equation (1):

$$F_i = x + (2i + 3/2i + 1)F_{i+1} \quad , \quad i = 0, \dots, n, \quad F_{n+1} = 0$$

According to the above formulas, the closer a sensor node is to the center, the more traffic will be loaded on it in the network.

3.1. The Scenario of Fuzzy logic system inference routing protocol

Fuzzy logic system is a generalization of the classical set theory. The most important advantage of fuzzy approach is that it is not necessary to have complete information and precise mathematical model of a system and therefore it is often easy to understand the underlying functionality of the system [15, 16]. In this scenario, selection of the next step to send the data packet by a node in the network is dependent on the following parameters:

- The most of the remaining energy (the first input of the Fuzzy logic system)
- Minimum traffic load (the second input of the Fuzzy logic system)
- The distance of the node (the third input of the Fuzzy logic system)
- The length of the queue (the fourth input of the Fuzzy logic system)

3.2 The systematic features of the proposed flat network

The Fuzzy logic system inference system is used to compute the cost of the node $NC_{(n)}$. This cost is defined by four parameters below:

- **The remaining energy:** The remaining energy of the neighboring node.
- **Traffic load:** This parameter is dependent on the number of data packets exchanged at the unit of time from node 'n'.
- **The distance from the node:** This parameter is derived from the distance between two nodes.
- **The length of the queue:** This parameter refers to the occupation of the node queue.

The Fuzzy logic system output that will be considered as the cost of the node or $NC_{(n)}$ is calculated by equation (2):

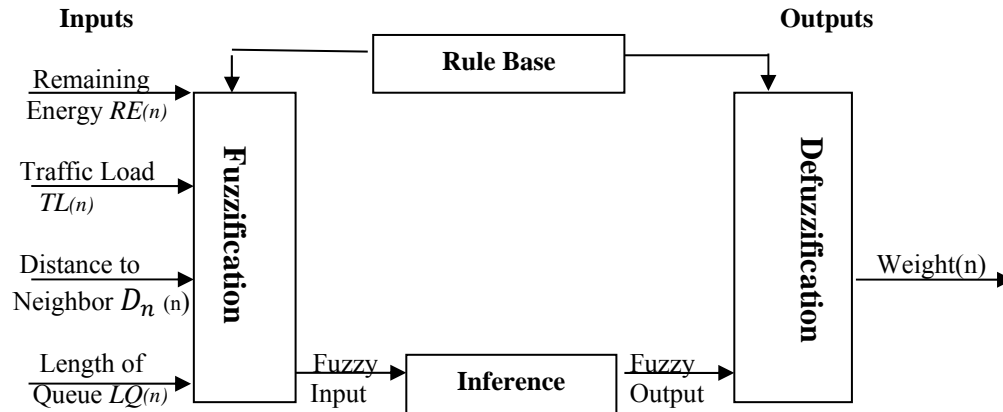
$$NC(n) = \sum_{i=1}^n U_i \times C_i / \sum_{i=1}^n U_i$$

Minimum step: The lowest step in the proposed algorithm named $MH_{(n)}$ is the shortest path that a node will use to send its data to the sink. The final function for this purpose took advantage of equation (3):

$$f(n) = NC(n) + 1/(MH(n))$$

Eventually, the sensor node with the highest value of $f(n)$ would be selected as the best node candidate to send packets to the sink. Thus, each source node within its communication range identifies its neighbors [17].

The proposed Fuzzy logic system has been designed according to the following figure.



(Fig 2: View of the proposed Fuzzy logic system with four inputs)

Each input parameter of the Fuzzy logic system Inference System consists of a triangulation graph. One of the diagrams relates to the remaining energy of the network. The energy of each node in the beginning of the network is assumed to be 10 J. This energy can be placed in five levels, which are composed of Very High, High, Medium, Low and Very Low. The higher the level of the remaining energy is the much higher priority the node has to be selected in the next step of a data packet [18]. The second input is the traffic load on a sensor node in the network. The higher the amount of traffic load on a node, the less priority it has to be selected for the next step.

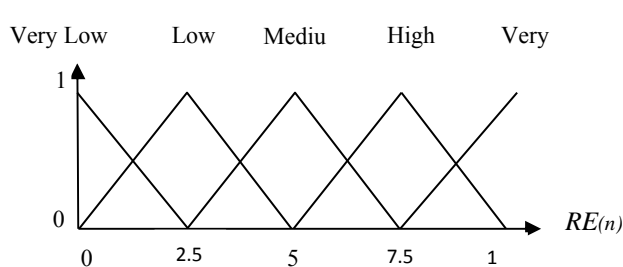


Figure 3: Remaining energy of node

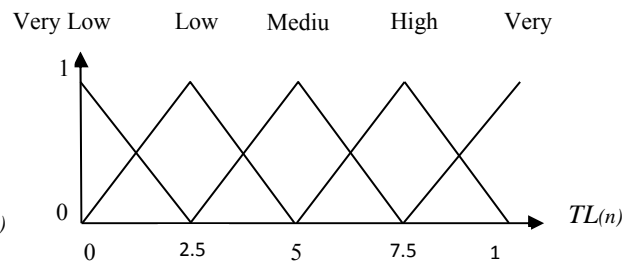


Figure 4: The traffic load of the node

The next input parameters are the distance of each node with its neighboring nodes and the length of queue which is assumed at five levels as displayed in the figure below.

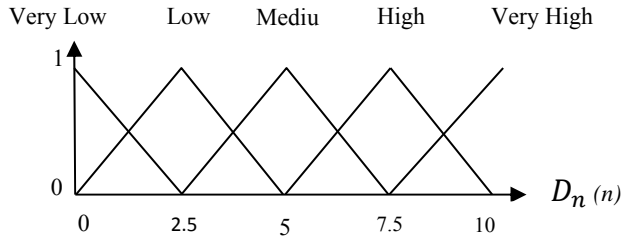


Figure 5: the distance from neighboring nodes

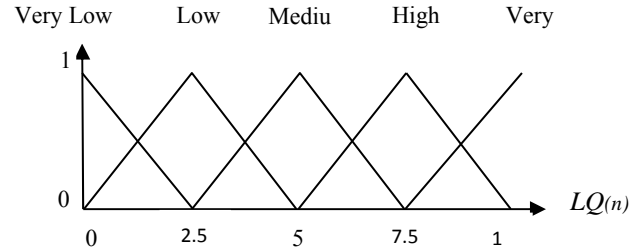


Figure 6: the queue length of nodes

According to equation (2) we found that the Fuzzy logic system input integrates into Fuzzy logic system diagram and the cost of each sensor node is calculated. In Figure 7, the cost of each sensor node is within $[0, 1]$. The higher the amount of this cost is the higher priority the node has in the selection. Finally, this output would become the input of the final function $f_{(n)}$.

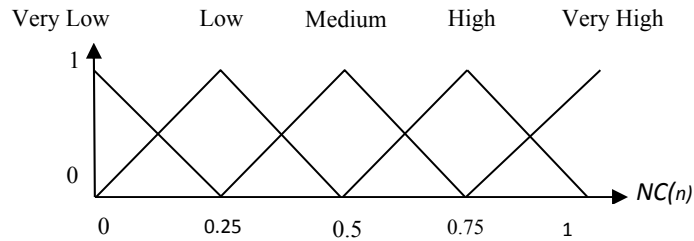


Fig 7: Node Cost

(Table1: Fuzzy rule base for fuzzy logic system to calculate the cost of each law)

<i>Antecedent</i>					<i>Consequent</i>
<i>Rules</i>	<i>Remaining Energy</i> $RE_{(n)}$	<i>Traffic Load</i> $TL_{(n)}$	<i>Distance to Neighbor</i> $D_{(n)}$	<i>Length of Queue()</i> $LQ_{(n)}$	C_i $NC_{(n)}$
<i>Rule1</i>	<i>Low</i>	<i>Medium</i>	<i>Very Low</i>	<i>Low</i>	<i>Medium</i>
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
<i>Rule16</i>	<i>Medium</i>	<i>Medium</i>	<i>Very Low</i>	<i>Medium</i>	<i>Medium</i>

3.3 Using learning automata for channel allocation

A learning automaton is a machine or decision-making unit to follow a predetermined sequence of operations or respond to encoded instructions [19, 20]. In the proposed method regarding how to allocate channels to nodes for sending the desired data some measures have been considered. As previously mentioned, the learning automata was used in cases of congestion. In the default mode, 6 radios are present of which 1 radio is used to send and receive routing control packets and 5 other channels are used to send and receive network data packets. In the first moment of the network, only two radios are active, because an increase in the number of radios causes sensor nodes to consume an increased amount of energy. As a result, we apply a threshold for congestion at the MAC layer which was considered in the previous section. Thus, as the congestion is increasing in the node, the number of active radios to send the package is gradually increasing. This increase in the number of radios with the probability of their use is done based on the learning automata.

In the first moment the random space of the automata has a higher initial probably to one of radios. The procedure for granting initial probability is random; and the rest of the radios receive equal percentages of the remaining probability. As an example:

Step 1: Radio number 2 has a prior probability of 40 percent. The rest of the radios have 15 percent probability in the current round.

$$S = \{0.4, 0.25, 0.25, 0.25, 0.25\}$$

After selecting channel 2, the sample is as follows:

$$S = \{0.35, 0.21, 0.21, 0.21\}$$

Step 2: In the event of congestion among the four remaining channels, we will randomly select one (the reward $a = 0.5$), and the probability of the channel is reduced or increased [21].

$$p_i^{(n+1)} = p_i^{(n)} + a[1 - p_i^{(n)}] \quad p_j^{(n+1)} = (1 - a)p_j^{(n)} \quad \forall j, j \neq i$$

$$p_{n+1} = 0.25 + 0.5 * (1 - 0.25) = 0.625 \quad \text{Increase the probability of channel}$$

$$p_{j(n+1)} = (1 - 0.5) * 0.25 = 0.125 \quad \text{Reduce the probability of other channel}$$

The sample space would be in this form:

$$S = \{0.625, 0.125, 0.125, 0.125\}$$

If all channels once are loaded and on, then the channel selection is not a random event, but it would be like this:

Favorable response from the environment: If the current channel congestion is less than the next channel, it will be considered a good one (a channel which has little congestion is appropriate for us and its probability is calculated by the former formula)

Step 3: The condition in which our automata will stop, is equality of the probability of all channels. Otherwise, the first stage will be repeated.

4. SIMULATION AND RESULT

(Table 2: Simulation Conditions)

<i>PARAMETERS</i>	<i>Queen-MAC ,TMCP</i>	<i>PROPOSED</i>
<i>Network size</i>	<i>1000*1000</i>	<i>1000*1000</i>
<i>Antenna</i>	<i>All the way</i>	<i>All the way</i>
<i>Time simulation</i>	<i>1000</i>	<i>1000</i>
<i>The number of sink</i>	<i>1</i>	<i>1</i>
<i>Location sink</i>	<i>Fixed</i>	<i>Fixed</i>
<i>Primary energy sink</i>	<i>1000</i>	<i>1000</i>
<i>The number of nodes</i>	<i>100</i>	<i>100</i>
<i>Position nodes</i>	<i>Random</i>	<i>Random</i>
<i>Routing Protocol</i>	<i>Flat ,Tree Based</i>	<i>Fixed</i>
<i>Primary energy</i>	<i>10 Jules</i>	<i>10 Jules</i>
<i>Energy model</i>	<i>Battery</i>	<i>Battery</i>
<i>Type Product</i>	<i>Environment Temperature</i>	<i>Environment Temperature</i>

4.1. Routing packet number test: This parameter indicates the number of packets that are sent and received, so that it can be concluded that how much the protocol has been able to deliver the produced packets in the network undamaged. Generally, the closer is the distance of the sent and received packets, the more efficient is the process of the protocol.

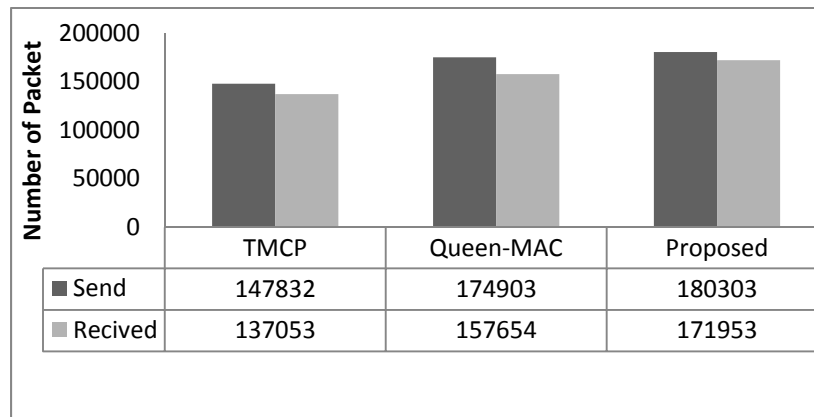


Fig 9: The diagram of the number of network passing packets

4.2. Data packets delivery rate test: This amount is on the percentage basis and is calculated based on the following formula. Evidently, the higher percentage indicates the better efficiency of the network.

$$PDR = \text{Data Received} / \text{Data Send} * 100$$

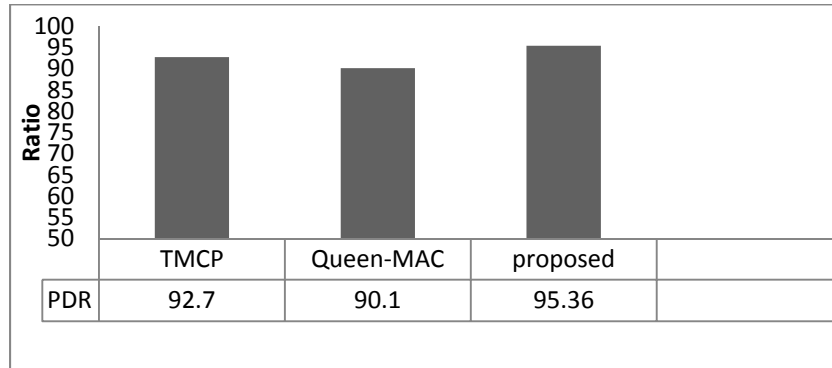


Fig 10: The diagram of the data packets delivery rate

4.3. The network remaining energy test: As the results indicate, at the end of the simulated time the proposed network still has 90% of the total primary. Whereas, similar protocols lost their energy quickly.

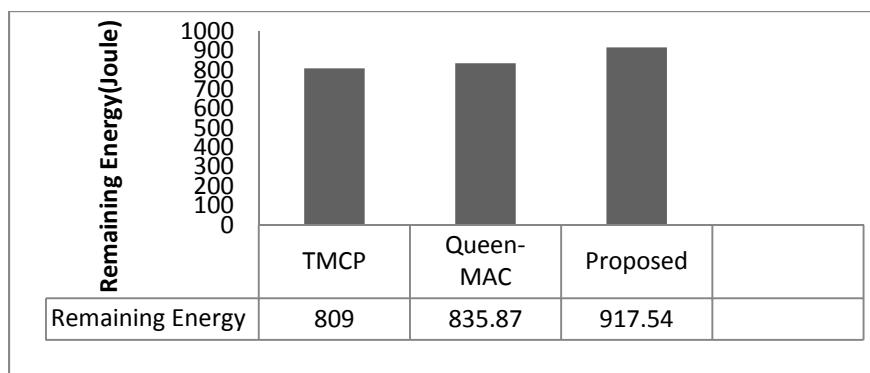


Fig 11: The diagram of the network remaining energy

4.4. The number of lost network packets test: The parameter to evaluate the efficiency of network with respect to routing can be the rate of the deleted packets in the network. Therefore, the lower number of removed packets indicates that the protocol correctly and in due time transferred the routing packets of the network.

$$Drop\ Packet = \sum Send - \sum Receive$$

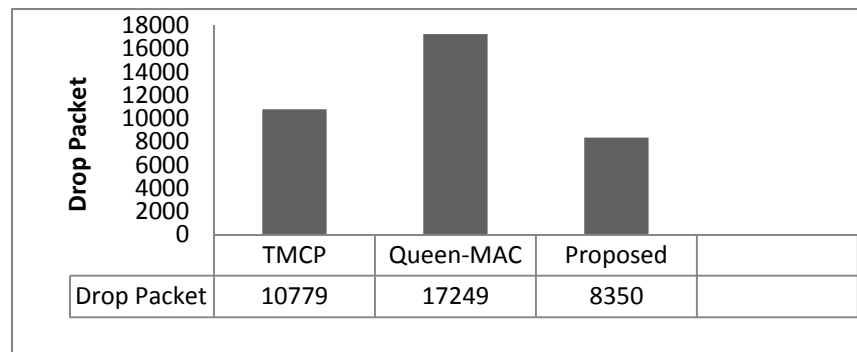


Fig 12: The diagram of drop packet

4.5. The tests of the effect of the number of circles (layers) on network performance

the impact of the number of hypothetical rings that exist as the category 0 to category 5 are compared in terms of the number of sent, received and deleted packets in the network. According to the output figures, it is indicated that the network with the number of 5-layer had a greater efficiency, and the network has been able to deliver more packets to the destination. In this test it was concluded that increasing the number of rings had a direct impact on their delay time to the end of the packets network. Therefore, the number of the lost packets will increase.

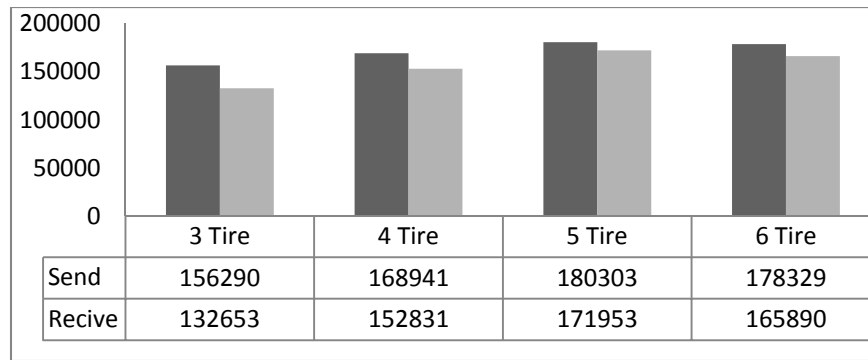


Fig 13: The diagram of packets received in the sink

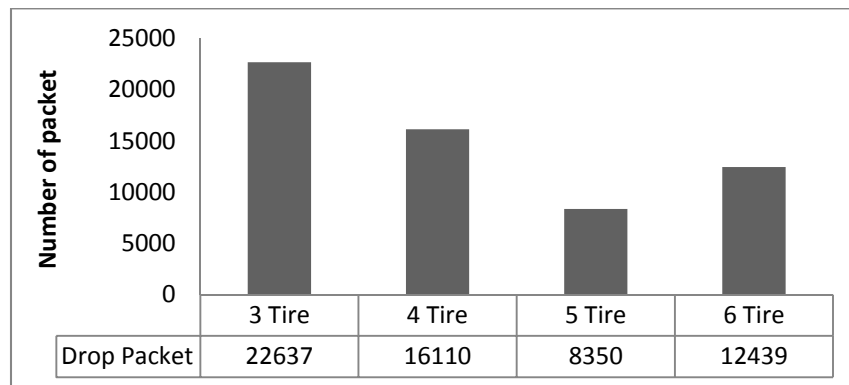


Fig 14: The diagram of drop packet in the sink

5. CONCLUSION

In this study, the method for preventing and dealing with congestion and loss of packets in the network is suggested. In this method, unlike the majority of similar protocols, a rapid and low-cost process was used to evaluate the qualifications of the routing nodes. The selection of the appropriate node in terms of the four remaining energy parameters, traffic load, the distance and the length of the queue of nodes can postpone the death time of the nodes and prevent simultaneous receipt of packets on a single node which can cause interference and ultimately the loss of packets.

REFERENCES

- [1]. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., & Cayirci, E. (2001). Broadband and Wireless Networking Laboratory. School of Electrical and Computer Engineering, Georgia Institute of Technology, 38, 393-422.
- [2]. Sakthidharan, G.R., & Chitra, S. (2012). A survey on wireless sensor network: An application perspective, In Computer Communication and Informatics (ICCCI).International Conference, 1-5.
- [3]. Ramakrishnan, M., & Vanaja Ranjan, P. (2009). MULTI CHANNEL MAC FOR WIRELESS SENSOR NETWORKS. International Journal of Computer Networks & Communications (IJCNC), 1, 2.
- [4]. Salarian, H., Chin, Kwan-Wu., & Naghdy, F. (2012). Coordination in wireless sensor-actuator networks. Journal of Parallel and Distributed Computing, 72,856-867.
- [5]. El-Hoiydi, A., & Decotignie, J.-D. (2005). Low power downlink MAC protocols for infrastructure wireless sensor networks. Mobile Networks and Applications, 10,675-690.
- [6]. Zheng, Jun., & Jamalipour, A. (2009). Wireless sensor networks: a networking perspective.
- [7]. Chao, Chih-Min., & Yi-Wei, Lee. (2010). A Quorum-Based Energy-Saving MAC Protocol Design for Wireless Sensor Networks. IEEE Transactions on Vehicular Technology, 59, 813-822.
- [8]. Chen, X., Han, P., He, Q.S., Tu, S.L., & Chen, Z.L. (2006). A multi-channel MAC protocol for wireless sensor networks. IEEE International Conference on Computer and Information Technology, 224-230.
- [9]. Chowdhury, K.R., Nandiraju, K.R., Chanda, P., Agrawal, D.P. & Zeng, Q.A. (2009). Channel allocation and medium access control for wireless sensor networks. Ad Hoc Networks, 7, 307-321.
- [10]. Yafeng, Wu., Stankovic, A., Tian, He., & Shan, Lin. (2008). Realistic and efficient multi-channel communications in wireless sensor networks. In Proceedings of the INFOCOM08, 27, 1193-1201.
- [11]. Gang, Zhou., Chengdu, Huang., Ting, Yan., Tian, He., Stankovic, A., & Abdelzaher Tarek, F. (2006). MMSN: multi frequency media access control for wireless sensor networks. In Proceedings of the INFOCOM06, 25, 1-13.
- [12]. Salajegheh, M., Hamed, S., & Antonis, K. (2007). Hymac: hybrid tdma/fdma medium access control protocol for wireless sensor networks. In Proceedings of the PIMRC07, 18, 1-5.
- [13]. Youngmin, K., Hyojeong, S., & Hojung, Cha. (2008). Y-MAC: an energy efficient multichannel MAC protocol for dense wireless sensor networks. In Proceedings of the IPSN, 7, 53-63.
- [14]. Ekbatanifard, Gh., Monsefi, R., Yaghmaee, M.H, & Hosseini, A. (2012). Queen-MAC: A quorum-based energy-efficient medium access control protocol for wireless sensor networks. Computer Networks, 56, 2221-2236.

- [15]. Esalat Nejad, A., Arbabi, M., & Romouzi, M. (2014). A Survey on fuzzy based Clustering Routing Protocols in Wireless Sensor Networks. *International Journal of Mechatronics Electrical and Computer Technology*.4, 2305-0543.
- [16]. Wang, L. (1996). *A course in fuzzy systems and control*. Prentice-Hall.
- [17]. AlShawi, Imad S., & Yan, Lianshan. (2012). Lifetime Enhancement in Wireless Sensor Networks Using Fuzzy logic system Approach and A-Star Algorithm. *IEEE SENSORS JOURNAL*, 12, 10.
- [18]. Zhang, H., & Shen, H. (2009). Balancing energy consumption to maximize network lifetime in data-gathering sensor networks. *IEEE Trans a Parallel Distrib*, 20, 1526–1539.
- [19]. Akbari Torkestani, J., & Meybodi, M. (2012). A learning automata-based heuristic algorithm for solving the minimum spanning tree problem in stochastic graphs. *J Supercomput* 59, 1035–1054.
- [20]. Narendra, K. S., & Thathachar, K. S. (1989). *Learning Automata: An Introduction*, New York, PrinticeHall.
- [21]. Shabanzadeh Dehkordi, S., & AkbariTorkestani, J. (2014). Channel Assignment Scheme based on Learning Automata for Clustered Wireless Ad-Hoc Networks. *Journal of Computer Science and Information Technology*. 2,149-172.

Presenting a Model to Meet the Pathology Component in the Implementation of Beyond Component Processes in Distributed Integrated Information Systems

Masoud Rafighi, Yaghoub Farjami
Department of Computer Engineering and Information Technology
University of Qom
Qom, Iran

Abstract- making all the applications in an enterprise work in an integrated manner, so as to provide unified and consistent data and functionality, is a difficult task because it involves integrating applications of various kinds, such as custom-built applications (C++/C#, Java/J2EE), packaged applications (CRM or ERP applications), and legacy applications (mainframe CICS or IMS). Furthermore, these applications may be dispersed geographically and run on various platforms. In addition, there may be a need for integrating applications that are outside the enterprise. According to the problems of adding application to organization and keep integration between them, in this paper, we studied the ways of integration between systems of organization. Then consider the Problems of models and emphasize on crucial need to create an ideal model for optimal architecture which meets the needs of the organization for flexibility, extensibility and integration of systems. Finally proposed a model which in addition doing comprehensive processes between the components easily in distributed systems, it does not have the problems of previous models. Since components are vulnerable in sending beyond component processes, so in this article we decided to introduce a model of pathology components to resolve the implementation of beyond component processes.

Keywords: ESB, Data-centric architecture, architecture Component-based, Plug in architecture, distributed systems.

I. Introduction

In addition to mainframe applications, which form the backbone of IT systems of large enterprises, the IT system of a large organization typically has a number of package applications. Examples of such package applications include Customer Relationship Management (CRM) applications and Enterprise Resource Planning (ERP) applications. SAP, PeopleSoft, Oracle, and JD Edwards are some of the software suppliers for these types of applications. Some of the advantages of these package applications for large organizations include risk reduction, introduction of best practices and processes, speed of implementation, and more accurate estimation of the cost of the software. Frequently these package applications are also referred to as Enterprise Information Systems (EISs). For such large organizations it is also important to integrate these package applications with the other applications in the IT systems in order to provide a consistent and unified view of data and functionality to both the internal and external customers.

Most, if not all, of the schemes for integrating these package applications rely on the use of adapters. Adapters are simply software components or subsystems that allow package applications to talk to other applications using the interfaces provided by the package applications. Modern ways of integrating these applications use adapters in conjunction with a J2EE application server to connect the EIS with the modern applications. Alternatively, the adapters can be used with an Enterprise Service Bus (ESB) to integrate the EIS with a wider variety of applications. In addition, sometimes the EIS supplier provides an infrastructure to expose some of the functionality and data embedded in the EIS application as Web Services.

II. Related Work

2.1 Different way of integration of application in Mainframe

Two broad categories of integration schemes are used in the point-to-point approaches when it comes to mainframe integration. The first broad category employs a messaging system for integrations. The second broad category of integration schemes exposes the mainframe applications as Web Services [1,23,19].

The first category consists of two methods:

- In the first method, the mainframe application is enabled to communicate directly with the messaging software system. The messaging software system then talks to the Java/J2EE application.
- In the second method, the messaging system does not talk directly to the mainframe application; instead, the connection is made through a bridge.

The second category also consists of two methods:

- In the first method, the mainframe application is directly exposed as a Web Service without the use of any middle service components. Only some versions of CICS can be exposed by using this method.
- In the second method, the mainframe application's functionality is first wrapped in a middle service component, which is then exposed as a Web Service [2,24].

2.2 The first way of using messaging system to integrate applications

In messaging, the applications do not communicate with each other directly and do not have a dedicated communication indirectly through queues. A queue-sometimes called a channel- behaves like a collection of messages that can be shared across multiple computers. In asynchronous messaging the code for the communication and marshaling is separated out as a separate software component, which allows for code reuse (that is, multiple applications can use the same code to communicate with each other and with applications on another machine). This separate software component is often called a messaging system or message-oriented middleware (MOM) [7,8,19].

2.2.1 The three elements of a basic messaging system are: [7,25].

❖ Channels or queues

Channels are used to transmit data, and each channel acts as a virtual pipe that connects a receiver with the sender.

❖ Messages

Messages encapsulate the data to be transmitted. A message consists of a header and a body. The information contained in the header is primarily for the messaging system to use. The header contains information regarding destination, origin, and more. The body contains the actual data the receiver consumes. The data contained in the body can be of different types. It can be a command message, which is used to invoke a procedure (method) in the receiving application, or it can be a document message, which is used to transfer data from one application to another. It can also be an event message, which is used to inform the receiving application of an event in the sending application.

❖ End points

A message end point contains a set of code that is used to connect to the messaging system and to send or receive a message. The rest of the application uses the end points whenever it needs to send or receive a message. Message end points are of two general types. The first type is used to send a message whereas the second type is used to receive messages.

Although distributed objects provided a big step forward on many fronts in the battle for enterprise applications integration, they failed to address two very important shortcomings of RPC:

- Both RPC and distributed objects employ synchronous interaction between the applications being integrated. This means that the client application is blocked from doing further work until the server application completes its work and returns control to the client application. This leads to strong coupling between applications and a lack of scalability in the integration solution. In other words, if a large number of applications need to be integrated, neither RPC nor distributed objects is the proper solution.

■ RPC- and ORB-based communication is not reliable and there is no guarantee that the messages and return values will be delivered to the intended targets. Thus, the client application may experience a hang-up in its operation under certain circumstances (such as a break in the network connection or when the two applications are not up and running at the same time) [9,19,26].

Asynchronous messaging overcomes these two major problems with RPC and distributed objects. In asynchronous messaging the client (or service consumer) sends a message to the server but does not wait for response. This allows the client application to perform further work while the server is completing the request from the client. This decoupling between the client and server means that more work can be accomplished in a given timeframe. In other words, it leads to a more scalable solution [4,5].

2.2.2 The benefits of messaging system

- Another important feature of a messaging system is that it can guarantee delivery of a message to the target application by persisting the message.
- Yet another important problem that asynchronous messaging solves relates to applications specifically designed to run disconnected from the network, yet synchronize with servers when a network connection is available. Examples include applications deployed on laptop computers and PDAs.
- Another important feature, which refers to the fact that with RPC and distributed objects, a single server can be overloaded with requests from different clients. This can lead to performance degradation and even cause the server to crash. Because the messaging system queues up requests until the server is ready to process them, the server can control the rate at which it operates on the requests so as not to overload itself by too many simultaneous requests [7,8,19].

2.2.3 Here are some of the disadvantages of asynchronous messaging:

- Generally speaking, asynchronous messaging software is costlier in monetary terms than the ORB-based middleware. For example, the cost of an ESB based on the asynchronous messaging middleware is typically more than ten times higher than the cost of an ESB based on an ORB-based middleware.
- A learning curve is associated with the asynchronous messaging environment.
- A certain amount of overhead and bookkeeping is involved in simulating a synchronous interaction between two applications [9,22].

2.3 Second way of using web services system to integrate applications

As mentioned previously, some of the package applications (EISs) directly expose some of their functionality and data as Web Services. For example, SAP directly exposes some of its functionality as Web Services. Any external application that has a network connection can use such functionality, thus providing another integration method for these package applications.

However, many times this direct exposure is not enough because the functionality needed by a consumer application may not be wholly contained in a single package application. In addition, because only some of the functionality of a given package application is exposed directly as Web Services, there is sometimes still a need to expose the remaining functionality as Web Services [27]. The method described previously in this chapter that employs adapters to integrate the package application with modern applications (particularly Java/J2EE applications) comes in handy. This is because once the functionality and data contained in the package applications have been integrated with J2EE components; it is easy to expose these components as Web Services [14,15].

2.3.1 The standards that Assigned to the web services include:

XML

XML is probably the most important pillar of Web Services. XML documents are often used as a means for passing information between the service provider and service consumer. XML also forms the basis for WSDL (Web Services Description Language), which is used to declare the interface that a Web Service exposes to the consumer of the service. Additionally, XML underlies the SOAP protocol for accessing a Web Service [10,15].

Web Services often pass information using XML documents. Therefore, the applications that implement Web Services or the applications that act as the consumer of Web Services must be able to interpret the information contained in an XML document. In addition, these applications must be able to extract and process the information contained in an XML document. Furthermore, they must be able to assemble XML documents from the results of this business processing [10,14].

SOAP

Simple Object Access Protocol (SOAP) is an XML-based messaging specification. It describes a message format and a set of serialization rules for data types, including structured types and arrays. This XML-based information can be used for exchanging structured and typed information between peers in a decentralized, distributed environment. In addition, SOAP describes the ways in which SOAP messages may be transported to realize various usage scenarios. In particular, it describes how to use Hypertext Transfer Protocol (HTTP) as a transport for such messages. SOAP messages are essentially service requests sent to some end point on a network. The end point may be implemented in a number of different ways, including an RPC server, a Java servlet, a Component Object Model (COM) object, and a Perl script, which may be running on any platform [11,14].

A SOAP message is fundamentally a one-way transmission between SOAP nodes, from a SOAP sender to a SOAP receiver. In other words, a SOAP message may pass through a number of intermediaries as it travels from the initial sender to the ultimate recipient [11,15,19].

WSDL

In order for a service consumer (application) to use the service provided by a service provider application, a formal description of the service is required that contains the description of the interface exposed by the service and information on where that service can be found on the network. Such a formal specification is provided by the Web Services Description Language (WSDL). A WSDL document is an XML-based document that describes a formal contract between the service provider and the service consumer [19,15].

A WSDL document describes two aspects of a service: the abstract interface exposed by the service and the description of the concrete implementation. The abstract interface describes the general interface structure, which includes the operations (that is, methods) in the service, the operation parameters, and abstract data types. The concrete implementation description binds the abstract interface description to a concrete network address, communication protocol, and concrete data structures. The concrete implementation description is used to bind to the service and invoke its various operations (methods) [12,14].

UDDI

In addition to the WSDL description of a service and the SOAP message format, a central place is needed where the service provider can advertise the services it offers and the service consumers can find the service they require. Such a central place is called a service registry [13,14].

WS-I Basic Profile

The Web Services Interoperability (WS-I) Organization is an open industry effort chartered to promote Web Services interoperability across platforms, applications, and programming languages. The organization brings together a diverse community of Web Services leaders to respond to customer needs by providing guidance, recommended practices, and supporting resources for developing interoperable Web Services. The WS-I Basic Profile provides constraints and clarifications to those base specifications (XML, SOAP, WSDL, and UDDI) with the intent to promote interoperability [16,17,19].

2.3.2 Advantages and disadvantages

The main objective standards related to Web services is presenting solutions for various problems of heterogeneity in large organizations.

Although these standards, known as Web Services, are able to solve some of the heterogeneity problems, they are not able to solve all of these types of problems. Some of the heterogeneity problems not addressed by Web Services standards include the following:

- ❖ Protocol mismatch Related to the heterogeneity of communication protocols is the problem that different applications want to communicate with each other using incompatible protocols. For example, Application A might want to communicate with Application B using HTTP. However, for Application B the suitable protocol might be IIOP. In such cases, a protocol transformation is needed so that Application A can communicate with Application B.
- ❖ Message format mismatch Related to protocol mismatch is the problem of a mismatch of message formats between the service provider and the service consumer. This problem refers to the situation where a service provider may be set up to receive messages in one format (such as SOAP), while the service consumer is set up to use another message format (such as Java RMI) [14,15].

2.4 BPMN 2.0

A standard Business Process Model and Notation (BPMN) will provide businesses with the capability of understanding their internal business procedures in a graphical notation and will give organizations the ability to communicate these procedures in a standard manner. Furthermore, the graphical notation will facilitate the understanding of the performance collaborations and business transactions between the organizations. This will ensure that businesses will understand themselves and participants in their business and will enable organizations to adjust to new internal and B2B business circumstances quickly [17,22].

2.5 Enterprise Service Bus (ESB)

The Enterprise Service Bus (ESB) pattern provides a comprehensive, scalable way to connect a large number of applications without the need for each pair of applications to make a direct connection [18,21]. Such a direct connection between two applications is called a point-to-point connection and the advantages of this indirect connection include:

- The number of connections needed is equal to the number of applications being integrated.
- Another great benefit of this indirect connection scheme through ESB is that it is easy to maintain and upgrade.
- Yet another advantage of this indirect connection scheme is that it provides more agility to the integrated structure [19,20].

Core functionalities

- Control and context-based routing
- Protocol transformation or switch
- Data or message transformation

With these three basic functionalities incorporated into an ESB's core, the ESB can offer a number of virtualizations [18]. The three main categories of virtualizations are as flow:

- Location and identity virtualization
The service consumer application does not need to know the address or location of the service provider application, and the service provider does not need to know the identity of the service consumer application. The service request can be filled by any one of a number of service providers [18,19]. This allows the service provider to be added or removed from the integrated structure without bringing down the system, thus providing for uninterrupted service to the service consumer
- Interaction protocol
The service consumer and service provider need not share the same communication protocol or interaction style.
- Interface

The service consumer need not agree on an exact match with the interface offered by the service provider. The ESB reconciles the difference by transforming the request message into the form expected by the service provider [19,22].

2.6 Pathology components include:

1. Pathology of implementation of beyond component processes.
2. Pathology of storing shared data in time of beyond component processes
3. Pathology of connecting components to each other
4. Pathology of sending data and operation in implementation of beyond component processes.

III. The proposed model

3.1 The proposed model

This architecture made of combining Data- centric architecture, plug-in architecture and component architecture so that in this architecture all components are connected to the data center but the components must appear with two hands (it is getting from plug-in architecture with this innovation that both hands SERVICE INTERFACE and Plug in interface added to every component. It means components have two hands instead of one hand). So in addition to connection they can transfer services and data. By using SOC discuss we concluded that every component must maintain its own data and just Common data such as Authentication and etc. will be kept in Data- Centric. We called the proposed architecture, CPDC Architecture which contains bellow parts:

- Data center: Information in the data center, public data, such as user categories, audition, authentication, access level and organizational chart of the organization need to be placed in the center.
- Service interface: An interface to transfer services from one component to another component
- Plug in interface: Certain protocol for connecting components
- Service: Services and operations that are performed on the data in each module
- Plug in manager: management, control and configure of plugin will done.
- Specific data: Data that is for a special system and there is no need to exist in other systems.
- Host component: The various modules which are available in the organization
- ESB: this part cause strengthened and better performance processes between components
- BPMN Engine: It Cause implementation process.

The Fig. 1 shows the proposed model. The following Fig. 2 shows that if the process be so long that we have cross several components to do it so how we can be done it with using the proposed model.

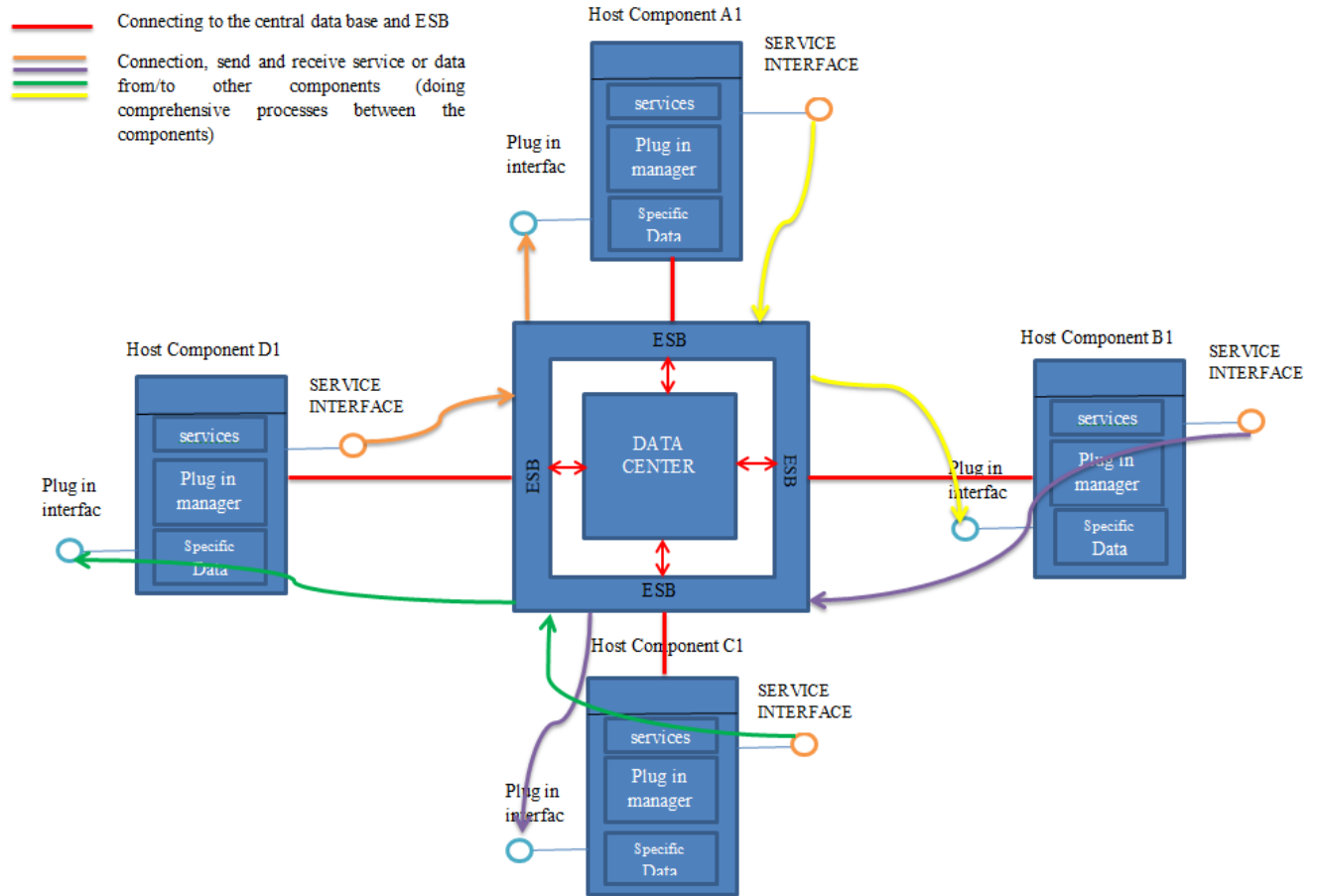


Figure 1 . The proposed model

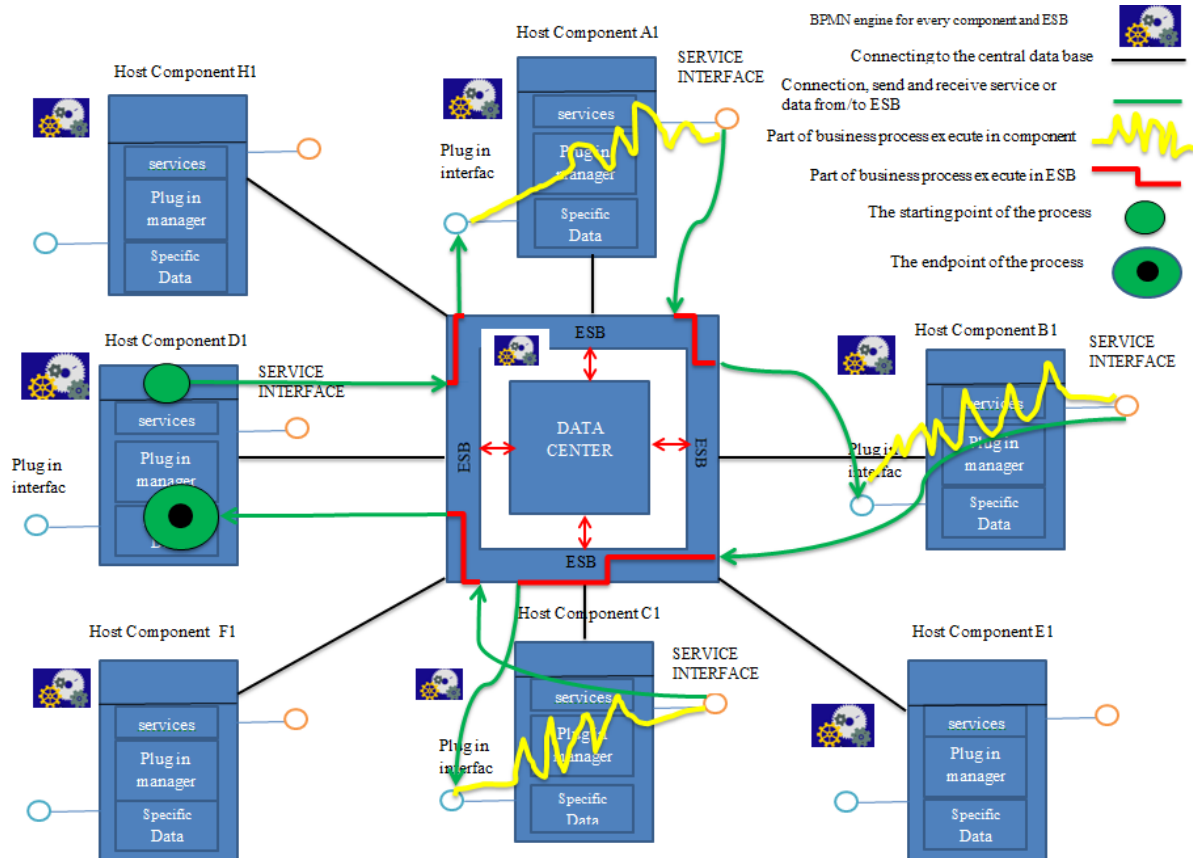


Figure 2 . doing comprehensive processes between the components

3.2 Differences between offered model and BPMS

Although the process flow with BPMS can be created and send operations, they use of one type of protocol within their own while any of the applications in integrated systems operate under certain protocol. The data format of sent is the same in a BPMS and does not need to convert Type but in integrated systems each of the applications use specific format to send and receive data. So it is clear that only with the help of a BPMS can't integrate organization's systems.

The following Fig. 3 shows the external view of processes between components shown in the proposed model.

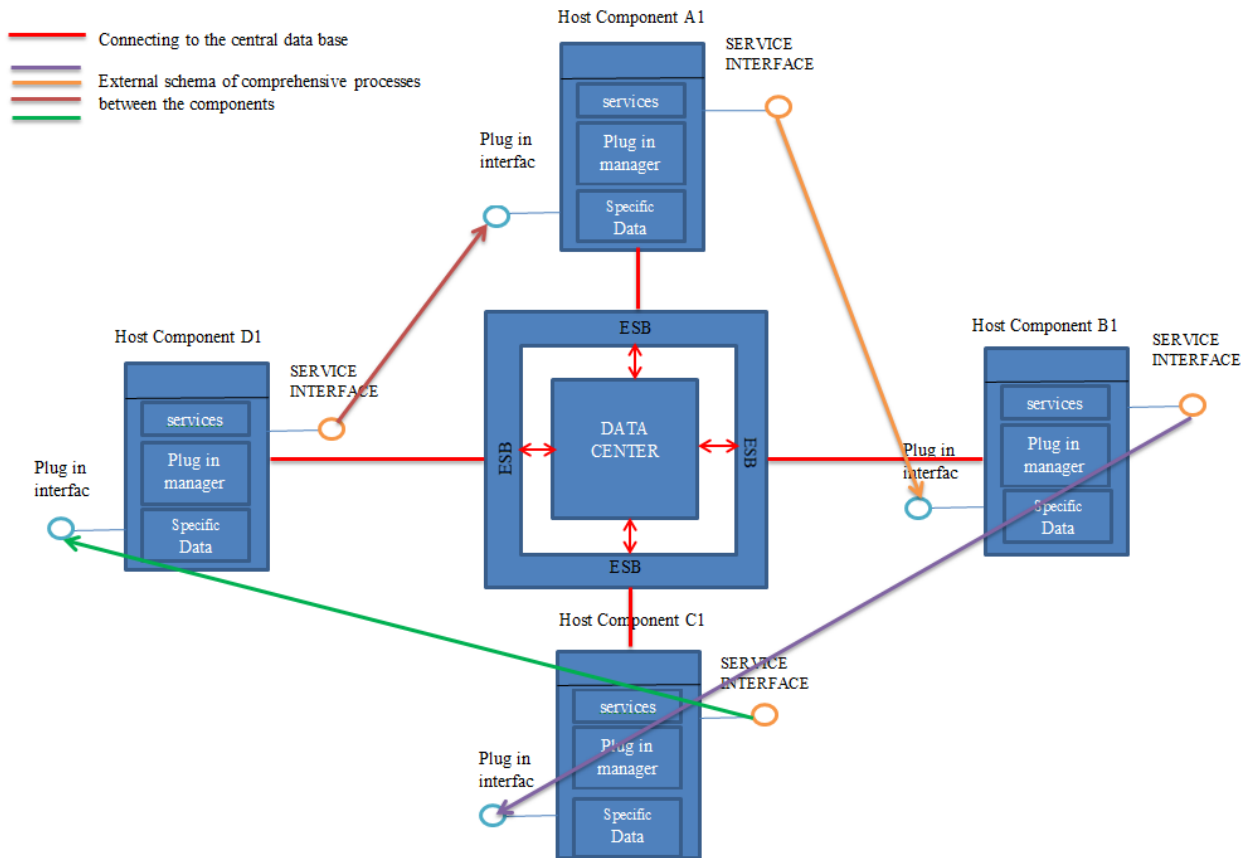


Figure 3 . External view of comprehensive processes between the components

IV. Conclusion

In proposed model components, the data and operation will be sent into ESB and they will be sent to other components by ESB. In fact common data of beyond component Processes will store in ESB memory. Given that organizational processes with internal processes components are different, in this way the owner of processes of organization (between the components) is central part of proposed model, the ESB. According to the existing the USB in model processes between components between older programs (which used old formats and protocols) and newer and more modern applications will connected as well. The proposed model includes the following features:

- Transforms between disparate message formats, including binary, legacy, and XML, and provides message routing and security, MQ/HTTP/FTP connectivity, and transport mediation.
- Provides transport-independent transformations between binary, flat-text, and other non-XML messages, including COBOL Copybook, ISO 8583, ASN.1, and EDI, to offer an innovative solution for security-rich XML enablement, enterprise message buses, and mainframe connectivity.
- Offers standards-based, centralized governance and security for proposed model, including support for a broad array of standards such as WS-Security and WS-Security Policy.
- Allows interaction among multiple heterogeneous applications, including native connectivity to registries and repositories, as well as direct-to-database access.

REFERENCES

- [1] Silberschatz, A., H.F. Korth, and S. Sudarshan. Database System Concepts, Fifth Edition, McGraw-Hill, 2005.
- [2] Elmasri, R. and S.B. Noble. Fundamentals of Database Systems, Second Edition, Addison-Wesley, 1994.
- [3] Folk, M.J., B. Zoellick, and G. Riccardi. File Structures: An Object-Oriented Approach with C++, Addison-Wesley, 1998.
- [4] Unix Network Programming: Interprocess Communication, Volume 2, Second Edition, Prentice-Hall, 1999.
- [5] Bloomer, John. Power Programming with RPC (Nutshell Handbooks), O'Reilly & Associates, Inc., 1992.
- [6] Henning, M. and S. Vinoski. Advanced CORBA Programming with C++, Addison-Wesley, 1999.
- [7] Hohpe, G. and B. Woolf. Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions, Addison-Wesley, 2004.
- [8] Yusuf, K. Enterprise Messaging Using JMS and IBM WebSphere, IBM Press, 2004.
- [9] Monson-Haefel, Richard. Enterprise Java Beans, Third Edition, O'Reilly, 2001 (message-driven beans).
- [10] Means W.S. and E.R. Harold. XML in a Nutshell: A Desktop Quick Reference, O'Reilly, 2001.
- [11] <http://schemas.xmlsoap.org/soap/envelope/> (SOAP schema).
- [12] <http://www.w3.org/TR/WSDL20/> (WSDL version 2.0).
- [13] <http://www.uddi.org/pubs/ProgrammersAPI-V2.04-Published20020719.htm> (UDDI API specification).
- [14] Zimmermann, O., M. Tomlinson and S. Peuser. Perspectives on Web Services: Applying SOAP, WSDL and UDDI to Real-World Projects, Springer, 2003.
- [15] Singh, I., S. Brydon, G. Murray, V. Ramachandran, T. Violleau, and B. Stearns. Designing Web Services with the J2EE 1.4 Platform: JAX-RPC, SOAP, and XML Technologies, Addison-Wesley, 2004.
- [16] <http://publib.boulder.ibm.com/infocenter/dzichelp/v2r2/index.jsp?topic=/com.ibm.etools.ims.tmra.doc/topics/tmresoverview.htm> (IMS, TM, Resource Adapter).
- [17] Juric, M.B., B. Mathew and P. Sarang. Business Process Execution Language for Web Services, Second Edition, Packt Publishing, 2006.
- [18] <http://www.redbooks.ibm.com/abstracts/sg247212.html> (Details of WESB, WESB red book).
- [19] Roshen W. "SOA-Based Enterprise Integration: A Step-by-Step Guide to Services-Based Application Integration" Copyright © 2009 by The McGraw-Hill Companies, 2009.
- [20] Hohpe, G. and B. Woolf. Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions, Addison-Wesley, 2007.
- [21] Fowler, M. Patterns of Enterprise Application Architecture, Addison-Wesley, 2002.
- [22] Manuel Oriol, Thomas Gamer, Thijmen de Gooijer, Michael Wahler, Ettore Ferranti, Fault-tolerant fault tolerance for component-based automation systems, in: Proceedings of the 4th International ACM SIGSOFT Symposium on Architecting Critical Systems (ISARCS 2013), Vancouver, Canada, 2013.
- [23] Masoud Rafighi, Yaghoob Farjami, Nasser Modiri, Assessing component based ERP architecture for developing organizations, International Journal of Computer Science and Information Security, vol-14-no-1-jan-2016, Pages 72-92.
- [24] William Otte, Aniruddha S. Gokhale, Douglas C. Schmidt, Efficient and deterministic application deployment in component-based enterprise distributed real-time and embedded systems, Inf. Softw. Technol. 55 (2)(2013) 475-488.
- [25] ERP SYSTEMS: PROBLEMS AND SOLUTION WITH SPECIAL REFERENCE TO SMALL & MEDIUM ENTERPRISES, Indu Saini, Dr. Ashu Khanna, Dr. Vivek Kumar, International Journal of Research in IT & Management, IJRIM, Volume 2, Issue 2 (February 2012).
- [26] SaaS Enterprise Resource Planning Systems: Challenges of their adoption in SMEs, Jacek Lewandowski, Adekemi O. Salako, Alexeis Garcia-Perez, IEEE 10th International Conference on e-Business Engineering, 2013.
- [27] Custom Development as an Alternative for ERP Adoption by SMEs: An Interpretive Case Study, Placide Poba-Nzaou & Louis Raymond, Information Systems Management, 02 Sep 2013. Published online: 21 Oct 2013.

AUTHORS



Masoud rafighi was born in Tehran, Iran on 1983/08/10. He is PHD student of Qom University. He receives M.Sc degree in computer engineering software from Azad University North Tehran Branch, Tehran, IRAN. He has recently been active in software engineering and has developed and taught various software related courses for the Institute and university for Advanced Technology, the University of Iran. His research interests are in software measurement, software complexity, requirement engineering, maintenance software, software security and formal methods of software development. He has written a book on software complexity engineering and published many papers.



Yaghoob Farjami received his PHD degree in Mathematics (with the highest honor) in 1998 from Sharif University of Technology, Tehran, Iran. He is Assistant Professor of Computer and Information Technology Department at University of Qom. His active fields of research are ERP, BI, and Information Security.

A NOVEL ENERGY EFFICIENT CONNECTED TARGET COVERAGE HEURISTIC IN WSN

Sunita Gupta, Ph.D. Scholar, Suresh Gyan Vihar University, Jaipur
Dr.K.C.Roy, Professor, Kautilya Institute of Technology & Engineering, Jaipur
Dr. Dinesh Goyal, Professor, Suresh Gyan Vihar University, Jaipur
Sakar Gupta, Associate Professor, Kautilya Institute of Technology & Engineering, Jaipur

Abstract: - Wireless Sensors Networks (WSNs) are able to work in insensitive environments where real observations by human being are dangerous, incompetent and sometimes not feasible. A most significant characteristic of a WSN application is lifetime. Wireless sensor network can be used till they can sense and communicate the sensed data to base station. Sensing as well as communication, both are important functions and they use energy. Energy management and scheduling of sensors can effectively help in rising the networks lifetime. Energy efficiency in a region monitored by a sensor network is achieved by dividing the sensors into cover sets. Every cover set is able to monitor the targets for a definite time period. At a time only single cover set is in active state and rest others are in low power sleep state. Thus energy is preserved and lifetime of Wireless Sensor Network is increased. Creating the greatest number of such set covers is proved to be an NPC problem. An energy minimization heuristic called Q-Coverage P-Connectivity Maximum Connected Set Cover (QC-PC-MCSC) is proposed. Functioning of Sensor nodes is scheduled in such a manner that they are having Q-Coverage and P-Connectivity constraint and thus they improves the working duration of Wireless Sensor Network. A comparative study of performance of QC-PC-MCSC and existing heuristic is also done over Energy Latency Density Design Space for Wireless Sensor Network.

Keywords:-Wireless Sensor Network, Connected Target Coverage, Network Lifetime, Cover Set, Coverage, Connectivity, Q-Coverage, P-Connectivity.

I. INTRODUCTION

Each sensor node in WSN is equipped with sensing, data processing and communication capabilities. The sensor nodes form a connected network and work collectively to accomplish the assigned tasks such as surveillance, environment monitoring and data gathering. Since sensors are low-cost devices, a large amount of sensors could be densely deployed inside or surrounding the interested phenomenon to provide the measurements with satisfactory accuracy. Generally, replacement of batteries is impractical. That's the reason for lifetime dependency of WSN on battery time. Energy-efficient algorithms are created for maximizing the lifetime of wireless sensor network. For proper data gaining in WSN, covering of all targets and Connectivity of sensors to base station, both are required. Also for the reliability purpose higher order of Coverage and Connectivity is required.

II. IMPORTANT QoS PARAMETERS IN WSN

Coverage is an elementary concern in a WSN. Coverage defines how a target is monitored by sensors [1, 2]. The sensing area of a sensor is a disk where the sensor is at the center. The radius of the sensor is called the Sensing Radius (R_s).

Three types of Coverage are there, called area Coverage, discrete points Coverage and barrier Coverage [1]. In case of area Coverage, the observation space is partitioned into smaller areas called fields [3]. Clearly, providing area Coverage is a sufficient condition for providing target Coverage, but may waste the precious battery energy. However, in barrier Coverage [4, 5] the sensing capability of a

sensor is presented as the likelihood that a sensor detects the phenomenon. Energy efficient management of resources and provision of reliable QoS are the two main needs in sensor networks [6]. Connectivity is the way to transfer the information sensed by sensor nodes to the base station. R_c is the Connectivity radius and a sensor may transfer the sensed data up to Connectivity radius. Q-Coverage and P-Connectivity are required for higher order of coverage and proper communication. Q-Coverage means every point of the plane is monitored by minimum q-different sensors [7] and P-Connectivity means a minimum of p disjoint paths should exist between any two sensors [7].

III. RELATED WORK

In [8], a centralized and node disjoint heuristic algorithm is proposed. It proficiently creates the cover sets that observe all the targets. Node disjoint means a sensor could take part in one cover set only. It is used for coverage and no concern about connectivity is given. In [9], authors proposed the solution of network life time maximization problem as the Maximum Set Covers (MSC) problem. It is centralized and node disjoint algorithm used for one as well as q-coverage algorithm. In [10], a heuristic is proposed called High Energy and Small Lifetime (HESL) and QoS requirement is added by considering Q-Coverage. In [11], authors introduce the Connected Set Covers (CSC) problem. It is centralized algorithm and both Coverage and Connectivity is considered in this algorithm. In [12] authors proposed a heuristic called Triple Phase Iterative Connected Set Cover (TPICSC) and it arrange the sensors into different subsets considering simple Coverage and Connectivity to the Base Station.

IV. DESCRIPTION OF PROPOSED HEURISTIC WITH Q-COVERAGE P-CONNECTIVITY MAXIMUM CONNECTED SET COVER (QC-PC-MCSC).

The parameters used in the propose heuristic are A, Q, P, l, E, e_1 and e_2 . Where A is the sensor target Coverage matrix. The value of A_{ij} is equal to 1 if a sensor S_i covers the target T_j . Otherwise it is 0. Q is the order of Coverage which is same for all sensors. P is the order of Connectivity as defined above. Each value of P-Connectivity vector is same here. l is a small sensor lifetime granularity constant, which is same for each set cover. E is the initial battery lifetime of each sensor. e_1 is the energy used for sensing and e_2 is the energy used for communication per unit of time.

Initially the battery lifetime of every sensor is set to E. Cover sets are made and all the four phases will be executed only if the Q-Coverage condition ($\sum_i A_{ij} B_i \geq q_j$) is satisfied. B is the Sensor Battery Life time set. k represents the number of set covers available. Initially the numbers of set covers (k) are set 0. The phases of the proposed heuristic are given below.

(1) Coverage Phase

In this phase the Coverage order is checked and a new set cover is formulated only if the condition of Q-Coverage is satisfied. A critical target is found out. For example the target most sparsely covered in terms of number of sensors as well as with regards to the residual battery of those sensors. After critical target selection, the heuristic find out the sensor having highest contribution or the sensor with the highest utility and that covers the critical target. The selected sensor is added to the current set cover. A target is either covered by the sensors already selected in the set cover, or it becomes a critical target, at which point the sensor with the greatest contribution, that covers the critical target, is selected again. This formulated cover set C_k will be used in next Connectivity phase.

(2) Connectivity Phase

In the Connectivity phase C_k , G and P are used as input. C_k is the set cover formulated in the above phase. G is the network Connectivity graph. P is the order of Connectivity. In this phase a new and updated connected set C_k is formulated.

BFS algorithm is used to find out the shortest path of each sensor S_i in C_k to the BS in G till the constraint of P-Connectivity is satisfied. Add all these selected relay nodes to the set cover C_k , forming an updated connected set C_k . The selected relay nodes are added to the set C_k to form an updated connected set C_k . So output of this phase is C_k and it is used in Redundancy Reduction phase

(3) Redundancy Reduction Phase

The aim of Redundancy Reduction Phase is to eliminate the maximum number of extra (redundant) sensors from C_k to minimize the number of sensors in the formulated connected set cover. Even if a sensor with higher order of Connectivity is redundant then a preferred sensors having a lower degree than this is selected for removal. Remove the sensor with minimum utility and then again check if it is still a connected set cover.

For this all sensors $S \in C_k$ are unmarked. Select the unmarked sensor $S \in C_k$ with the least degree in C_k or minimum utility. Then it is checked if the set $C_k - \{S_i\}$ is still a connected set i.e. it ensures connected Coverage in the Wireless Sensor Network. If it is true, then modify the set C_k as $C_k = C_k - \{S_i\}$, Else mark the sensor S_i . This process is repeated till all $S \in C_k$ are marked. Thus output of this phase is updated set C_k , which will be used in Life Time Assignment and Energy Updation Phase.

(4) Life Time Assignment and Energy Updation Phase

A small lifetime is assigned to the set cover C_k , generated in Redundancy Reduction Phase. The time period of a cover set is determined as minimum between small lifetime granularity constant (l) and maximum lifetime available from sensors in a set cover C_k . Thus every cover set is active for l_k time. The energy consumed by an active sensor for sensing is equal to $E_1 = l_k e_1$, and for communication is equal to $E_2 = l_k e_2$ in a round.

So an active sensing sensor uses $E_1 + E_2$ energy. An active relay sensor use only E_2 energy per round. After the updates, if the remaining energy B_i of a sensor S_i is less than E_2 , then that sensor is eliminated from the set S .

V. QC-PC-MCSC HEURISTIC

INPUT (A, Q, P, l, E, e1, e2)

Lifetime of every sensor is set to E.

$k=0$

Repeat all 3 steps while $\sum_i A_{ij} B_i \geq q_j$ is true for each target

(1) Coverage Phase

$k=k+1$

$C_k = \phi$

For every targets

Uncover_level (T) = q_j

Do while uncover_level (T) $\neq 0$, for all targets

A critical target is selected with Uncover_level (T) > 0 and a sensor S that have greatest contribution function.

$C_k = C_k \cup \{S\}$

For every targets that are covered by S

Uncover_level (T) = Uncover_level (T) - 1

End do

(2) Connectivity Phase

Repeat for 1 to P

BFS algorithm is executed to find out the shortest path from each $S \in C_k$ to BS in G .

Extra nodes in this path are added to C_k to form a new and updated connected set C_k

End for

(3) Redundancy Reduction Phase

Unmark all $S \in C_k$

Repeat while all $S \in C_k$ are marked.

Choose unmark $S \in C_k$ with minimum utility.

If $C_k - \{S\}$ is still a connected set then $C_k = C_k - S$

Else mark that S

End while

(4) Lifetime Assignment and Energy Updation Phase

$l_k = \text{Lifetime}(C_k) = \min(l, \text{Max_lifetime}(C_k))$

For all Sensors $S \in C_k$

If S_i is performing as only relay node

Then $B_i = B_i - E_2$

Else if S_i is performing as sensing node then

$B_i = B_i - (E_1 + E_2)$

Else if $B_i < E_2$ then

$S = S - S_i$

End for

VI. SIMULATION OF QC-PC-MCSC

A sensing region of 1000x1000m is taken for the simulation of QC-PC-MCSC. The proposed heuristic is implemented using MATLAB and results are analyzed. All sensors contain equal energy, sensing radius and communication radius. For simulation, sensor and target numbers are taken in the interval [20,150] and [20, 90] respectively. As shown below in figure 1, the graph is drawn between the targets and lifetime for fixed number of sensors. The graph is drawn for fixed $q_m = 2$, $p_m = 1$. Different values of l is considered in the graph.

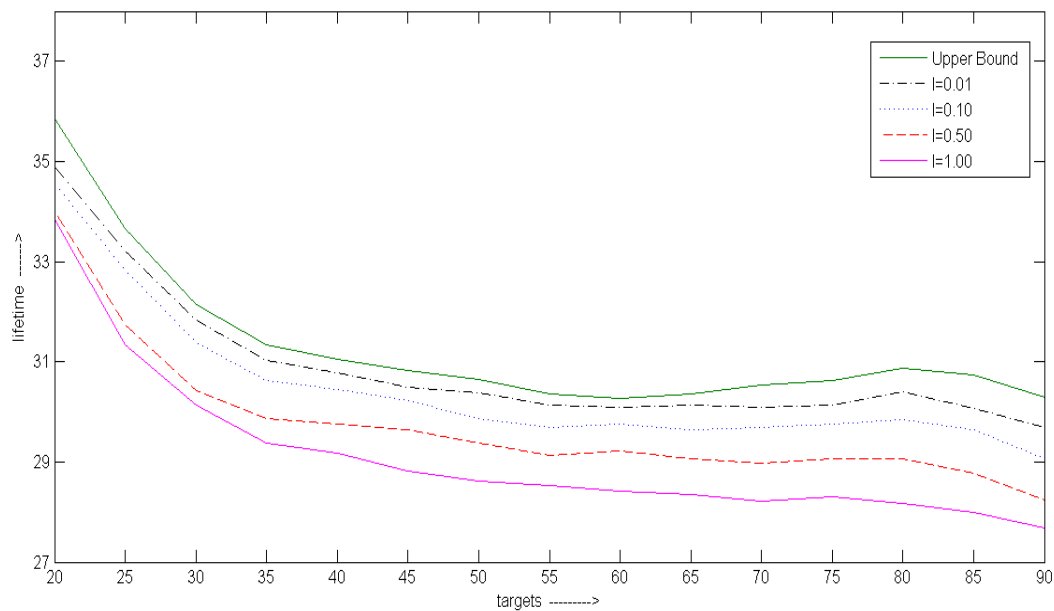


Figure 1:- The Lifetime Obtained by QC-PC-MCSC for $q_m = 2$, $p_m = 1$ and for Different Values of Targets.

Figure 2 contains the graph between the sensors and lifetime for fixed number of targets. Graph is drawn for $q_m = 2$, $p_m = 1$. Different values of l are considered in the graph.

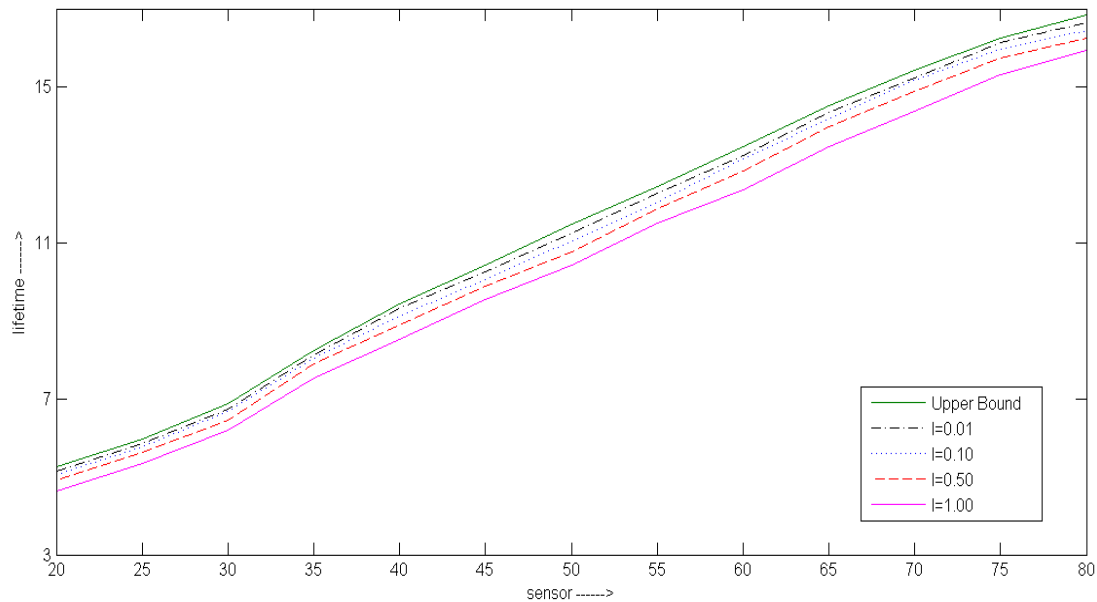


Figure 2:- The Lifetime Obtained by QC-PC-MCSC for $q_m=2$, $p_m=1$ and for Different Values of Sensors.

VII. Comparative Performance of QC-PC-MCSC and Other Existing Heuristics over Energy Latency Density Design Space Model

The model given in [13] is used to evaluate the performance of proposed QC-PC-MCSC heuristic and existing heuristic [10].

Energy Latency Density Design Space is a topology management application that is power efficient designed by Joseph Polastre, Jason Hill, David Culler [13]. A mathematical model of the network is designed with required energy, latency and density configuration using the model proposed in [13], to analyze the Performance of Proposed heuristic and existing heuristic in terms of time of a sensor.

The energy of a node is calculated by the overall lifetime of the nodes such as in [13]. The node's lifetime is inversely proportional to total energy consumption. Total energy E consumed by a node is sum of the energy used in receiving (E_{rx}), transmitting (E_{tx}), listening for messages on the radio channel (E_{listen}), sampling data (E_d) and sleeping (E_{sleep}). The notations and values used are listed in Table 1. Total energy used is given by

$$E = E_{rx} + E_{tx} + E_{listen} + E_d + E_{sleep} \quad (1)$$

All parameters used for the energy consumption E are same as given in [13]. The energy consumed in sampling data E_d , is

$$E_d = t_d C_{data} V \quad (2)$$

Where, $t_d = t_{data} \times r$

t_d is the time of sampling data, t_{data} is the sample sensors, r is the sample rate (packets/s), C_{data} is the current of sample sensors (mA), V is the voltage.

$$E_{tx} = t_{tx} C_{txb} V \quad (3)$$

Where, $t_{tx} = r \times (L_{preamble} + L_{packet}) t_{txb}$

t_{tx} is the time to switch the transmitter, $L_{preamble}$ is the preamble length (bytes), L_{packet} is the packet length (bytes), t_{txb} is the time (s) to transmit 1 byte, C_{txb} is the current required to transmit 1 byte, V is the supply voltage.

$$E_{rx} = t_{rx} C_{rxb} V \quad (4)$$

Where, $t_{rx} \leq nr (L_{preamble} + L_{packet}) t_{rxb}$

t_{rx} is the time (s) to switch the receiver, n is the neighborhood size of the node, t_{rxb} is the time (s) to receive 1 byte data, C_{rxb} is the current required to receive 1 byte data.

Table 1: Parameters Used for Calculations of Energy Consumption

Variables	Parameter	Values
C_{sleep}	Sleep Current (mA)	0.033
C_{batt}	Capacity of battery (mAh)	2600
V	Voltage	3.0
$L_{preamble}$	Preamble Length (bytes)	271
L_{packet}	Packet Length (bytes)	36
t_i	Radio Sampling Interval (s)	100E-3
R	Sample Rate (packets/s)	1/300
L	Expected Lifetime (s)	-

The low power listening check interval called LPL interval, should be less than the time of the preamble,

$$L_{preamble} \geq [t_i / t_{rxb}]$$

The power used in a single LPL radio sample is taken as 17.3μJ. The total energy used in listening the channel is the energy of a single channel sample multiplied by the channel sampling frequency.

$$E_{sample} = 17.3\mu J$$

$$t_{listen} = (t_{rinit} + t_{ron} + t_{rx/tx} + t_{sr}) * 1/t_i \quad (5)$$

$$E_{listen} \leq E_{sample} * 1/t_i$$

Where, t_{rinit} is the initialize radio time, t_{ron} is the turn in radio time, $t_{rx/tx}$ is switch to rx / tx time, t_{sr} is the time to sample radio.

The node must sleep for the rest of the time. So sleep time t_{sleep} , is given by

$$t_{sleep} = 1 - t_{rx} - t_{tx} - t_d - t_{listen}$$

and

$$E_{\text{sleep}} = t_{\text{sleep}} C_{\text{sleep}} V \quad (6)$$

The lifetime of the node (T) depends on the capacity of the battery (C_{batt}) and the total energy consumed by the battery (E) and is given by:-

$$E = \frac{C_{\text{batt}} \times V}{T} \quad (7)$$

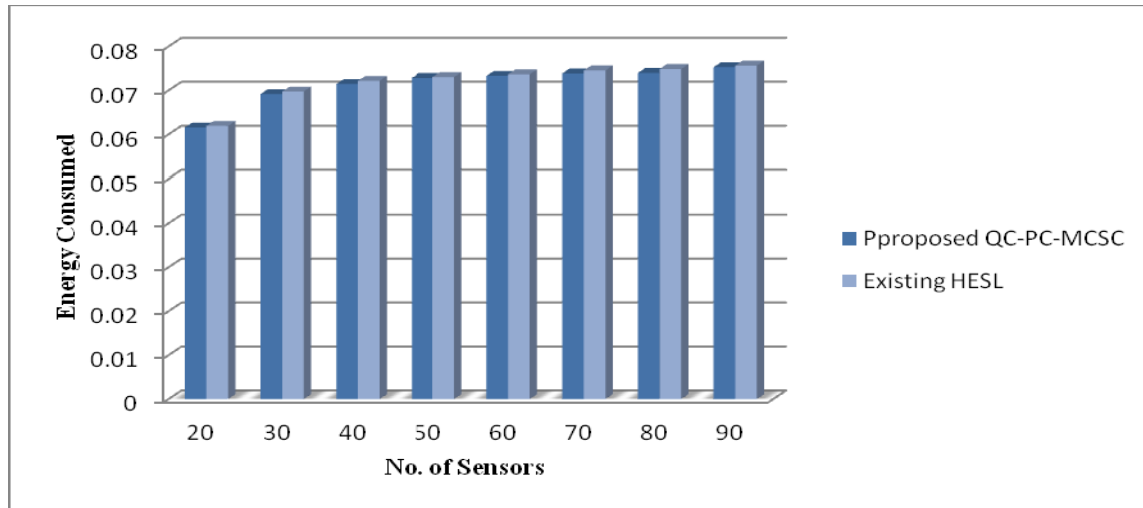


Figure 3:- Comparative Performance Analysis of QC-PC-MCSC Heuristic with Existing Heuristic over Energy Latency Density Design Space

The mathematical model designed for evaluation of performance quantifies the QC-PC-MCSC heuristic. The above calculations, results and graphs prove that the energy consumption by the sensors in the QC-PC-MCSC heuristics is less as compare to existing heuristic.

The real time implementation of the QC-PC-MCSC heuristic may help in implementing low cost Wireless Sensor Networks with high efficiency. A comparative performance of QC-PC-MCSC and existing heuristics is done over Energy Latency Density Design Space for Wireless Sensor Network.

VIII. CONCLUSION

A centralized heuristic for Q-coverage and P-connectivity problem is given in this paper. Simulations and analysis of result are done using MATLAB. The results show that the proposed method gives solution which is very near to the optimal solution. QC-PC-MCSC uses the greedy approach. A comparative performance of QC-PC-MCSC and existing heuristics is done over Energy Latency Density Design Space for Wireless Sensor Network which is a topology management application that is power efficient. The same problem can be varied by having additional constraints of Coverage and Connectivity or directional sensing etc.

IX. REFERENCES

- [1] Mohammad Ilyas and Imad Mahgoub eds., "Coverage Problems in Wireless Ad Hoc Sensor Networks", Handbook of Sensor Networks, chapter 19, CRC Press, 2004.
- [2] C.F. Huang and Y.C. Tseng, "A Survey of Solutions to the Coverage Problems in Wireless Sensor Networks", Journal of Internet Technology, vol. 6, no. 1, pp.1-8, 2005.
- [3] S. Slijepcevic and M. Potkonjak, "Power efficient organization of Wireless Sensor Networks", in Proc. of International Conference on Communications (ICC'01). IEEE, June 2001, pp. 472-476.
- [4] M. S., K. F., P. M. and S. M., "Coverage Problems in Wireless Ad-Hoc Sensor Networks," in Proc. of Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM), 2001, pp. 1380-1387.
- [5] M. S., K. F., Q. G. and P. M., "Exposure in Wireless Ad Hoc Sensor Networks," in Proc. of ACM International Conference on Mobile Computing and Networking (MOBICOM), 2001, pp. 139-150.
- [6] Amitabha Ghosha, Sajal K. Dasb, "Coverage and Connectivity issues in Wireless Sensor Networks", Pervasive and Mobile Computing 4 (2008) 303-334.
- [7] X. Bai, S. Kumar, D. Xuan, Z. Yun and T. H. Lai. "Deploying Wireless Sensors to Achieve Both Coverage and Connectivity". In Proc. of ACM MobiHoc, 2006.
- [8] Dimitrios Zorbas, Dimitris Glynos, Panayiotis Kotzanikolaou Christos Douligeris, "An Adaptive Algorithm for Coverage Problem in Wireless Sensor Networks", "IEEE INFOCOM'05 March".
- [9] Mihaela Cardei, My T. Thai, Yingshu Li, Weili Wu, "Energy Efficient Target Coverage Problems in Wireless Sensors Networks", "IEEE INFOCOM 2005, 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE, Page(s): 1976 - 1984 vol.3.
- [10] Manju Chaudhary, Arun K Pujari, "Q-Coverage Problem in Wireless Sensor Networks", International conference on Distributed Computing and Networking (ICDCN), 2009, Springer publications.
- [11] Ionut Cardei, Mihaela Cardei, "Energy Efficient Connected Coverage in Wireless Sensor Network", International journal of sensor networks, Volume-3, issue-3, May 2008, Pages 201-210, ACM digital library.

- [12] Jamali, M.a., Bakhshivand, N. , Easmaeilpour, M. , Salami, D.,” An Energy-Efficient Algorithm for Connected Target Coverage Problem in Wireless Sensor Networks”, Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference ,Volume: 9,Page(s): 249 - 254 , Publication Year: 2010 ,IEEE Conference Publications.
- [13] Joseph Polastre, Jason Hill, David Culler, ‘Versatile Low Power Media Access for Wireless Sensor Networks’, November 3–5, 2004.

A Novel Hybrid Encryption Scheme to Ensure Hadoop Based Cloud Data Security

Danish Shehzad¹, Zakir Khan², Hasan Dağ³, Zeki Bozkuş⁴

^{1,4}Department of Computer Engineering, ³Department of Management Information Systems,
Kadir Has University, Istanbul, Turkey

²Department of Information Technology, Hazara University, Mansehra, Pakistan

Abstract - Cloud computing and big data have provided a solution for storing and processing large amount of complex data. Despite the fact that they are quite useful, the threat to data security in cloud has become a matter of great concern. The security weakness in Hadoop, which is an open source framework for big data and cloud computing, has setback its deployment in many operational areas. Different symmetric, asymmetric, and hybrid encryption schemes have been applied on Hadoop for achieving suitable level of data security. In this paper a novel hybrid encryption scheme, which combines symmetric key algorithm using images as secret keys and asymmetric data key encryption using RSA, is proposed. The suggested scheme reduced the overhead of the secret key computation cycles as compared to the other existing encryption schemes. Thus, it is safe to claim that the proposed scheme retains adequate security level and makes data encryption more efficient.

Keywords: Hadoop, Hadoop distributed file systems (HDFS), Matlab, Data encryption scheme (DES), RSA.

I. INTRODUCTION

Hadoop is an open source platform developed under the Apache license for storing and processing large amounts of data [1]. It scales from single node to thousands of nodes for provisioning of storage and parallel processing capacity. Hadoop is based on two main modules: Mapreduce for processing and generating large data sets and Hadoop Distributed File System (HDFS) for storing data on distributed clusters [2,3]. Hadoop has been commonly accepted in the field of cloud computing where resource utilization and system performance require an excellent task scheduling mechanism [4]. Many users share the same resources, but the most critical issue in Hadoop environment is the data security, which is the main concern for improving the trust and dependability of organizations. Academic areas and industrial spheres have started their dependency on cloud computing. Due to the continuous availability requirement and various cloud applications running in parallel, it has become hard to achieve a high level of data security [5]. Cloud computing environments require data security at each and every level.

Different cryptographic techniques (symmetric and asymmetric) have been applied to encrypt data so that protected data is received at destination [6]. If private key cryptography is intended to be adopted by the two parties, then a secret key will be shared. Therefore, secure communication is limited between those who have pair of trusted keys. The major drawback is how securely key is transferred. Public key cryptography is used to solve the aforementioned problem. Hard mathematical problems use public key for encryption and private key for decryption on both sender and receiver sides [7]. Moreover, two parties want a guaranteed approach about the confidentiality and integrity of data. To resolve these issues feasible solutions and mechanism should be adopted. The concerns about RSA algorithm whether it is feasible to work with encrypted data or not, without having to decrypt the data first were revealed in [8]. This concern resulted in the research for hybrid encryption systems, which can combine symmetric and asymmetric encryption schemes, and it started real efforts for development of new hybrid encryption schemes. The basic requirement for this type of system can be understood as clients have confidential information, which they

send to servers for some computation without giving private key to the servers. So, hybrid encryption methodology can be used to obtain adequate level of security [9].

Strong encryption algorithms should be designed because computational power of machines is increasing day by day. Therefore, hybrid model provides better non linearity to plain RSA[9]. The likelihood of algebraic attacks on hybrid encryption models have recently increased, but on the other hand, the combination of RSA with DES resulted in better diffusion. The use of hybrid encryption scheme require more computations as compared to individual DES or RSA implementation, so as a result, hybrid encryption scheme consumes additional encryption time than the time required for individual DES and RSA implementations. It is no doubt that one needs to develop an alternate model that can reduce encryption time but provides the highest level of data security. Symmetric key algorithm using images as secret keys instead of DES reduced the secret key computation cycles overhead in Triple encryption mechanism retaining its hybrid nature and making it more efficient and secure [10].

The rest of the paper has been organized in sections, such as; section 2, which elaborates the literature review, section 3, which explains the proposed encryption scheme including symmetric key implementation using images as secret keys, after which data key encryption is done with RSA, and section 4, which gives the experimental results between symmetric key algorithm using images as secret keys and other existing symmetric encryption algorithms. Finally, section 5 provides some conclusions and elaborates the future research direction.

II. THE RELATED WORK

Data protection has become one of the main research topics in a cloud computing environment. Data security is the critical issue in cloud distributed data storage systems [5]. The detailed analysis of privacy threat in cloud scenarios was done by Siani Pearson, et al. [11]. They explained that security concern varies based on area of application. Strong mutual authentication using Kerberos was presented in [17], whereas central server is responsible for access control to storage servers, since data confidentiality on servers can be broken when storage servers are compromised by attackers. Various public key encryption schemes for cloud computing on Hadoop have been proposed. Giuseppe Ateniese et al. proposed data forwarding functionality scheme based on proxy re-encryption mechanism [13]. Another method using secure virtual machine was proposed by HouQinghua et al. [14]. The method concentrates on securing privacy of user data on cloud storage.

Yu Shu-cheng et al. proposed a technique of attribute based encryption for access control and data security in cloud computing [15]. Another method having a master - slave architecture for secure distributed file system was presented by Tahoe [16]. This technique used the Advanced Encryption Standard (AES) where data encryption/decryption keys are managed by the owner. Recently Tahoe's algorithm has been integrated to Hadoop for improving the security of data in Hadoop echo system. The task of key management increases and computation becomes heavier as each file needs different key and the owner has to manage key diversity along with increase in number of files. HDFS stores files in clear text and control the file security through a central server [13]. Therefore, HDFS security is considered as weak in Hadoop context, the communication between data nodes and clients and specifically among data nodes is not encrypted. Triple Encryption Scheme was proposed as hybrid encryption scheme based on DES algorithm as symmetric key encryption algorithm and RSA as public key encryption scheme, but the overall computation overhead is bottleneck for this hybrid mechanism [10].

III. THE PROPOSED SOLUTION

Data Hybrid Encryption using Image Secret Keys and RSA

In this scheme HDFS files are encrypted using hybrid encryption scheme in which files are encrypted symmetrically by image secret key and then this key is asymmetrically encrypted using owner's public key. User keeps the private key and sends the encrypted file to HDFS. Hybrid encryption combines the beneficial features of both symmetric and asymmetric encryption schemes. This hybrid encryption mechanism is shown in Fig.1.

Fig. 1. Hybrid encryption mechanism.

i. File Encryption and Data Key Generation

In this mechanism specified images, whose pixel values can represent all characters are selected for file encryption. The probability of characters in image is checked before using it for key generation. There can be multiple existences of a character, but any character is selected randomly from them. If any character is missing in the image one of the two options is used : either image is rejected or image is modified by inserting pixel having the specified character into the image. These images are selected as encryption key by data key management module and message letters are converted into corresponding 8-bit binary codes, which are scanned for image pixel values in image. When an appropriate match is found between pixel values code and message 8-bit code the location of the pixel is saved, where the locations are saved column wise. In the next step based on user ID, the file encryption/decryption module locates user public key based on RSA to encrypt data key and stores the encrypted key in database, as shown in Fig 2.

Fig.2. Flow chart for file encryption using image
as secret key.

Fig.3. Flowchart for file decryption using image
as secret key.

ii. File Decryption and Data Key Acquisition

When a user requests to download the file, it calls API to get file from Hadoop distributed file system to the application server, in response data key management module is requested to get the data key. Based on the file id, server queries the database where encrypted data key is located, then private key is used for restoring the data key, which is an image file, is returned by encryption/decryption module to decrypt the file. The flow chart for decryption process is shown in Fig.3.

IV. THE EXPERIMENTAL RESULTS

For the proposed hybrid scheme, initially image based symmetric key encryption is implemented in MATLAB to calculate computational cost of the proposed scheme and compared to the existing symmetric key encryption schemes. The experiments were conducted in Matlab (R2014b, 8.4 version) on Lenovo B5400, Core i5 having 2.6 GHz processor, 4GB RAM and 64 bit Windows 7 operating system. Each experiment was performed 5 times for obtaining result, and the results for each reading is the average of 5 runs as shown in Table 1. Various files of different sizes were taken and encrypted to measure the computational cost. Encryption time of any scheme is time taken to convert plain text to cipher text and the throughput of an encryption algorithm is calculated by dividing the total number of bytes in plain text by total encryption time. Table 1 shows execution time comparison between the proposed image based symmetric key encryption and the existing encryption schemes.

Table 1. Comparison of execution time(ms): The proposed technique vs the existing symmetric encryption schemes using variable data size

Input Size (KB)	DES	3DES	AES	Proposed Technique
102	83	112	122	7.38
124	72	108	102	8.64
200	104	142	154	14.8
500	122	178	189	37.48
640	152	246	298	50.72
1392	257	362	344	106.34
1788	384	466	431	133.6
1922	398	458	381	152.4
10698	2058	2289	1884	802.44
14628	2644	2996	2302	1094.3
Throughput (MB/sec)	5.099458	4.348783	5.154503	13.28599311

Fig. 4. compares throughput performance of the proposed and that of the existing encryption schemes. As seen from the figure clearly the performance of the proposed scheme is better than that of AES, 3DES, and DES. If overall average of encryption and decryption of the proposed technique is considered, it is more efficient than all the existing symmetric key encryption schemes. When the proposed symmetric key encryption is be combined with RSA it will ensure adequate security level along with a proficient key computation mechanism.

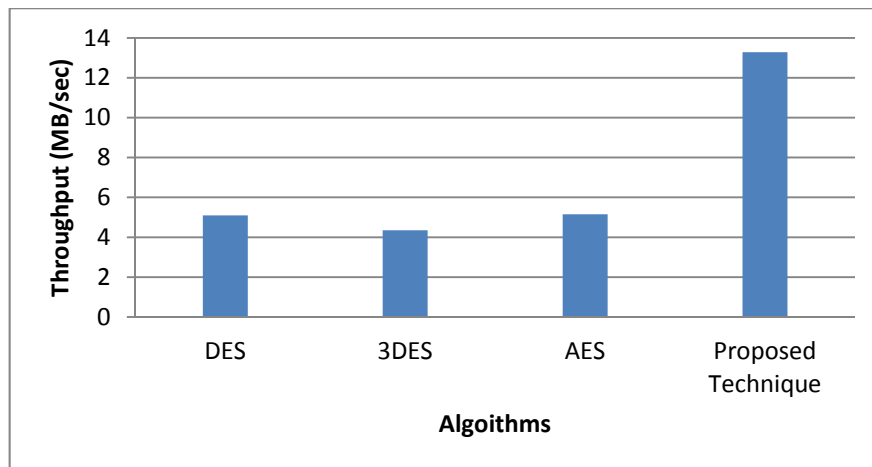


Fig.4. Throughput comparison of symmetric encryption algorithms.

V. CONCLUSION

A novel hybrid key encryption scheme is proposed to ensure proper Hadoop based cloud data security. The proposed scheme is based on hybrid mechanism, which uses image as secret key for symmetric encryption. The private key is used with RSA to generate public key. Initially the symmetric encryption algorithm is implemented in Matlab to ensure data security and to analyze its performance. The results shown appreciable increase in throughput for the proposed scheme over the existing symmetric key encryption schemes. In the future studies, the RSA would be implemented along with image based symmetric key encryption to ensure better security level and to evaluate the results to ensure efficient Hadoop based cloud data security.

REFERENCES

- [1] *Hadoop Distributed File System*, <http://hadoop.apache.org/hdfs>.
- [2] White, T, *Hadoop: The Definitive Guide*. O'Reilly Media, 2009.
- [3] J. Dean and S. Ghemawat, *MapReduce: Simplified data processing on large clusters*, in Proc. 6th Symp. Operating System Design and Implementation (OSDI'04), pp. 137–150, Dec. 2004.
- [4] P. Zikopoulos and C. Eaton, *Understanding Big Data: Analytics for enterprise class Hadoop and streaming data*, McGraw-Hill OsborneMedia, 2011.
- [5] Jam, M.R.; Khanli, L.M.; Javan, M.S.; Akbari, M.K., *A survey on security of Hadoop*, in the proceedings of the 4th International eConference on Computer and Knowledge Engineering (ICCKE), 2014, vol., no., pp.716-721, 29-30 Oct. 2014.
- [6] Katz, J. and Y. Lindell, *Introduction to modern cryptography*. Second edition 2014: CRC Press.
- [7] Hellman, M.E., *DES will be totally insecure within ten years*. *IEEE Spectrum*, 1979. **16**(7): p. 32-40.
- [8] Rivest, R.L., A. Shamir, and L. Adleman, *A method for obtaining digital signatures and public-key cryptosystems*. *Communications of the ACM*, 1978. **21**(2): p. 120-126.
- [9] Lin, H.-Y., et al. *Toward data confidentiality via integrating hybrid encryption schemes and Hadoop Distributed File System*. in the Proceedings of the 2012 IEEE 26th International Conference on Advanced Information Networking and Applications (AINA), pp. Pages 740-747.
- [10] Yang, C., W. Lin, and M. Liu. *A Novel Triple Encryption Scheme for Hadoop-Based Cloud Data Security*. *Fourth International Conference on Emerging Intelligent Data and Web Technologies (EIDWT)*. Xi'an, China 2013. IEEE.
- [11] Pearson, S. *Taking account of privacy when designing cloud computing services*. in *Proceedings of the 2009 ICSE Workshop on Software Engineering Challenges of Cloud Computing*. 2009. IEEE Computer Society.
- [12] O'Malley, O., et al., *Hadoop security design*. Yahoo, Inc., Tech. Rep, 2009.
- [13] Ateniese, G., et al., *Improved proxy re-encryption schemes with applications to secure distributed storage*. *ACM Transactions on Information and System Security (TISSEC)*, 2006. **9**(1): p. 1-30.
- [14] Qinghua, H., et al., *A method on protection of user data privacy in cloud storage platform*. *Journal of computer research and development*, 2011. **48**(7): p. 1146-1154.
- [15] YU Shu-cheng, Cong Wang, Kui Ren, Wenjing Lou. *Achieving Secure, Scalable, and Fine-grained Data Access Control in Cloud Computing*, in Proceedings of the 29th conference on Information communications, IEEE Press Piscataway, NJ, USA, 2010, Pages 534-542.
- [16] Wilcox-O'Hearn, Z. and B. Warner. *Tahoe: the least-authority filesystem*. in the Proceedings of the 4th ACM international workshop on Storage security and survivability. 2008.
- [17] Owen O'Malley, Kan Zhang, Sanjay Radia, Ram Marti, and Christopher Harrell. *Hadoop security design*. <https://issues.apache.org/jira/secure/attachment/12428537/securitydesign.pdf>, October 2009.

Enhancing Users' Satisfaction Using an Agent-Based Negotiation System

Omid R. B. Speily¹, Yosra Bahrani², Negin Razavi Rajayi²

¹ Department of Information Technology & Computer Engineering, Urmia University of Technology, Urmia, Iran.

² Department of Information Technology & Computer Engineering, AmirKabir University of Technology, Tehran, Iran.

Abstract-The increasing importance of operating automated systems arises with emerging competitive e-commerce environment. Nowadays, operating automated systems used in conducting all business transactions are enhanced substantially to achieve beneficial trade and decrease frequent messaging overhead of transactions. In spite of the highly competitive electronic marketplace, it is necessary to design a system which automates tasks including group negotiation and, payment and delivery. In this paper, we apply the purchasing groups to enhance the bargaining power of customers still satisfying all users' needs and preferences. We propose a flexible system called UUT-Trade to purchase laptop computers. This system uses a novel negotiation algorithm which diminishes all prices offered by potential sellers as much as possible, and then users will have the chance to choose between potential sellers by performing a weighted voting. Unlike similar systems which also exploit group purchasing, this system suggests no scarification of buyers' needs.

Key words: Negotiation, Automation, Scarification, UUT-Trade, AHP tree.

I. INTRODUCTION

The increasing importance of operating automated systems arises with emerging competitive e-commerce environment. In this paper, we will propose a system developed in a multi agent framework which uses the C2B e-commerce model to conduct its business transactions[1]. It is a flexible system which exactly satisfies user's preferences and tries to enhance the bargaining power of customers by forming purchasing groups. The proposed system exploits a new algorithm to bargain with potential sellers whom address preferences of a group. Our system uses AHP (Analytical Hierarchy Process) tree [2]. The Analytic Hierarchy Process (AHP) is an approach to multiple criteria decision making developed by Thomas Saaty in the early 1970 [3]. In [4] the AHP is defined as a theory of measurement concerned with deriving dominance priorities from paired comparisons of homogeneous elements with respect to a common criterion or attribute. In [5] an adaptive recommender system is introduced based on AHP to represent user's preferences and help him/her to find the best products in an electronic catalogue. In our proposed system AHP is used to synthesize customers' opinions about the weights of descriptive parameters being used to evaluate the quality of a product, unlike the other proposed systems (e.g. [6]) which synthesize customers' needs. In other words, our proposed system uses AHP tree as a tool, for measuring quality of a product. This allows gaining better understanding of users' preferences without being scarified[7]. The main reason of the scarification introduced in the authors' BCP model ([6]), is using AHP tree to synthesize users' preferences which will naturally scarify some of them. This system gives the most possible liberty to all users so that they can express exactly what they would like to buy. Using AHP tree for synthesizing users' needs will naturally scarify some preferences to get a consensus on purchasing ([6]). We propose a flexible system, called UUT-Trade to purchase laptop computers. The functional steps of this system will be discussed in section 2.2 to give you a better view on all operational aspects of the system.

During this paper, all issues of importance related to UUT-Trade system such as System Architecture, Bargaining Algorithm, Message Structure of the system, are tried to be completely covered and addressed. During this paper we explain how we improve the BCP model proposed by [6], and how we use collective purchasing while we satisfy all users' preferences. In other words, through UUT-Trade System, we explain how it is possible to use BCP model to increase the bargaining power of customers and still satisfy all of them. By UUT-Trade, we show general ideas related to an agent based negotiation system (conducted in Electronic Marketplaces), in a more intuitive way. Unlike [6], whom consider all components of a laptop (in their prototype) as a negotiable issue, we feel that, after a product has been produced, that it is unreasonable to request a seller to change one or more features of his supplied product (e.g. the CPU model of a laptop computer) to acquire a better deal, because it may not be possible for all sellers (for example if we are negotiating with retailers rather than manufacturers) to do this. That is, the second drawback of the system proposed by [6], is the limitation on negotiation with manufacturers, only. You will not be able to negotiate with retailers using their proposed system, whom also address the preferences of a group. This limitation will naturally decrease the bargaining power of customers, because negotiation will be just performed in the more limited community of manufacturers. So from negotiation perspective, we consider each product as being in a two dimensional space. Dimension (1): Price, and dimension (2): Descriptive Parameters Score (DPS). Our proposed negotiation (or bargaining) mechanism tries to buy products with lower prices and higher qualities (which is described by Price and DPS, respectively, in our system), while

it also gives the chance to users to vote on the negotiated products to express their desires, which is made possible by performing a weighted voting. Besides, we have predicted a Recommender System [8][9] in UUT-Trade, to provide recommendations for the users, whom are interested in joining purchasing groups, which will guide them in finding the most proper purchasing group to join. As also mentioned before, this system is implemented in a multi-agent framework, but we have not already introduced automated negotiation agents and negotiation itself. In [10] and [11] negotiation is defined as: "Negotiation (or bargaining) is the interaction that occurs when two or more parties attempt to agree on a mutually acceptable outcome in a situation where their orders of preference for possible outcomes are negatively correlated". In [12] intelligent software agents is defined as "programs to prepare bids for and evaluate offers on behalf of the parties they represent with the aim of obtaining the maximum benefit for their users". Authors in [13] have defined the agents from perspective of consumer buying behavior, as being engaged in the following activities: need identification, product brokering, buyer coalition formation, merchant brokering, and negotiation.

The remainder of this chapter is organized as follows. Section 2 is about related studies about negotiation systems. Section 3 discusses about our proposed system named UUT-trade system and its architecture. Evaluation of our system is discussed in section 4 and the last section discusses about conclusion and future works.

II. RELATED WORK

At the first, decision support systems are used by negotiators, but the necessity for negotiation support systems is detected in 1970s and after that various types of these systems are designed and developed to facilitate and automate the negotiation activities [14]. In this section, some of these systems are introduced.

In [15], a global multi criteria decision support system named Web-HIPRE has been introduced. Web-HIPRE helps people for individual and group decision making. This software is available from everywhere by locating on the WWW [16]. In [17], Decisionarium has been explained that is a web based software that uses Web-HIPRE and some other similar tools for interactive multicriteria decision support. In [13], authors has surveyed the state of agent mediated e-commerce, while they specifically concentrated on the B2C and B2B aspects. They also discussed the roles of agents in B2B e-commerce. They did this through B2B transaction model which identifies agents as being responsible for partnership formation, brokering, and negotiation activities.

Regarding Intelligent Software Agents, [18], has proposed a general agent architecture. This proposed architecture linked aspects of perception, interpretation of natural language, learning and decision-making. [19][20], has considered the negotiation task as an optimization problem and solved that by their proposed approaches. They has assumed that participants are given their individual profit schedules and each of participants desires to maximize his own profit obtained. [21], [22] and [23], has proposed the automated agent based negotiation in the electronic marketplace. They has assumed that the profit schedules are not given and only the offers by the participants are available, which seems to be more realistic than the assumption taken by [19] [20]. [21], has presented a two-fold agent based system with each part supporting interactive recommendation and automated negotiation activities. His proposed system supports activities which were most related to the decision making process. [24], has used web services and intelligent agent techniques to design a distributed service discovery and negotiation system, which were supposed to be operated in B2B ecommerce model. He also has developed an integrative negotiation mechanism to conduct multi-party multi-issue negotiations. He also has conducted an empirical study to evaluate his intelligent agent-based negotiation mechanism and to compare the negotiation performance of his software agents with that of their human counterparts.

In [6], a Buyer Collective Purchasing (BCP) model has been developed and implemented in a multi-agent framework for C2B e-commerce. They has addressed (1) how to synthesize individual's preferences into a group's consensus, (2) how to communicate with each other within the group using automated agents, and (3) how to collectively negotiate with a seller, etc., in their proposed BCP model. They also has developed a prototype system to show general ideas and how their proposed model works. As also the authors ([6]) approve, their proposed model (BCP) has drawback of scarifying users' needs and preferences to get a consensus, but our system fixes this problem, because we synthesize users' opinions on the quality of a product, unlike the authors ([6]) whom synthesize users' preferences. Synthesizing users' preferences will naturally scarify some preferences in order to get a consensus ([6]).

In [12], a multiple-attributes 4-phase negotiation model (information collection, search, negotiation, and evaluation) has been presented for B2C e-commerce. In this model, intelligent agents were deployed to facilitate autonomous and automatic on-line buying and selling. They also has applied fuzzy theory and analytical hierarchy process (AHP) to develop the system interface to facilitate the user inputs. They has assumed that buyer agents and seller agents have their own negotiation strategy, and developed a new negotiation strategy to obtain new offers gained by potential sellers.

In [25], a multi-agent model is developed that uses big data and business analytics to help sellers predict the buyers' negotiation strategy. In this model, buyer information is stored in system and based on analytics results, agents are able to negotiate with several relevant sellers and present the best offers to the buyer. Therefore this model improves the quality of negotiation decisions for both seller and buyer.

In [26], an agent-based approach has been proposed to multiple resource negotiation. This method uses case-based reasoning to select efficient sellers and resources and also learning automata is used for choosing the best negotiation strategy. This approach causes enhancement in some performance measures.

III. PROPOSED METHOD

This section discusses our proposed system in a step by step approach. The system contains two major conceptual steps, including primary step and secondary step, with sub-steps describing the main tasks accomplished by this system. Before starting to explain the steps we are going to explain System Architecture of the system. The System Architecture of UUT-Trade is explained in the section III.2.

1. System Architecture

As also mentioned in section 1, the UUT-Trade will operate in a multi-agent framework which supports the tasks performed by the system. The system employs 8 types of intelligent software agents. Functional role of each agent and messages exchanged between them are presented in fig. 1. In return, a brief explanation on the role of each agent is presented:

- **Group Former Agent:** As the name suggests, this agent is responsible for forming new groups and managing all existing groups. When a user wants to join to an existing group or wishes to create a new group, he/she connects to this agent and requests information about existing groups by his/her agent. This agent is kind of *the manager of all system components*, and in fact, acts as the server which serves to buyer agents. All information about existing groups will be maintained in a database only accessible by the Group Former Agent.
- **Domain Expert Agent:** The responsibility of gathering information about product's evaluation criteria can be performed by domain experts or an intelligent agent. In the proposed system, domain expert agent is engaged to update the weightless AHP tree but weightless AHP is designed manually by domain specialists. Due to the complexity of forming weightless AHP, automated agents should be carefully designed to do this task as well as knowledgeable specialists. After this agent has finished its task, a *Weightless AHP Tree* will be constructed. After achieving AHP tree, it should be delivered to the Group Agent.
- **Group Agent:** Each group is provided with a specific agent called Group Agent. This agent is responsible for weighting the Weightless (and probably updated) AHP tree obtained from Domain Expert Agent. This agent is exactly responsible for gathering users' opinions about the weights of AHP tree branches and, synthesizing them and finally obtaining Synthesized AHP tree. Synthesized AHP Tree should be delivered to Negotiator Agent[21][22].
- **Search Agent:** This agent is responsible for searching products of different companies (potential sellers) to find the matching products to the Group Specifications. This agent surfs the internet to find some products, which exactly satisfy Group Specifications and then in return, will communicate the results to the Negotiator Agent.
- **Negotiator Agent:** At the first place, this agent calculates Pr/DPS ratio (this ratio will be discussed in return) for each product based on the *Search results* and *Synthesized AHP Tree*, which are obtained from the search agent and Group Agent. The calculated values then act as the criteria on which the Negotiation will be accomplished. This Agent Negotiates with potential sellers, bargains with them and then communicates the winner seller to Contractor Agent.
- **Contractor Agent:** Eventually, this agent is responsible for Forming a Contract between Buyers and Winner Seller. This contract must be signed by the winner and by individual group members. This Agent is also responsible for settling payments. Hereafter, the products are ready for delivery.

2. Functional Steps Of UUT-Trade

There are one primary step and 9 secondary steps. These steps are presented in the following:

Primary step: Collecting descriptive parameters for the related product and making weightless AHP tree (by a specified automated agent or by domain experts). (Constructing)

Secondary steps:

1. Forming purchase groups and developing them. (Forming)
2. Updating the weightless AHP if necessary. (Updating)
3. Making weighted AHP tree *which describes the score of different sellers' products from a specified buyer's perspective*, and synthesizes individual Weighted AHP Trees to obtain *Synthesized AHP Tree*. (Synthesizing)
4. Searching for potential sellers, selling the products which match to the preferences of each group. (Searching)
5. Communicating search results to individual members of each group and gaining their approval. (Communicating)

6. Calculating “price” to “descriptive parameters score (DPS)” ratio (Pr/DPS) for each selected products. (Calculating)
7. Running negotiation agent and making potential sellers compete with each other. (negotiating)
8. Performing a weighted voting to select between the bargained products. (Voting)
9. Contracting.

The subsequent sub-sections of section 2 investigate briefly the overall steps, with details of each step.

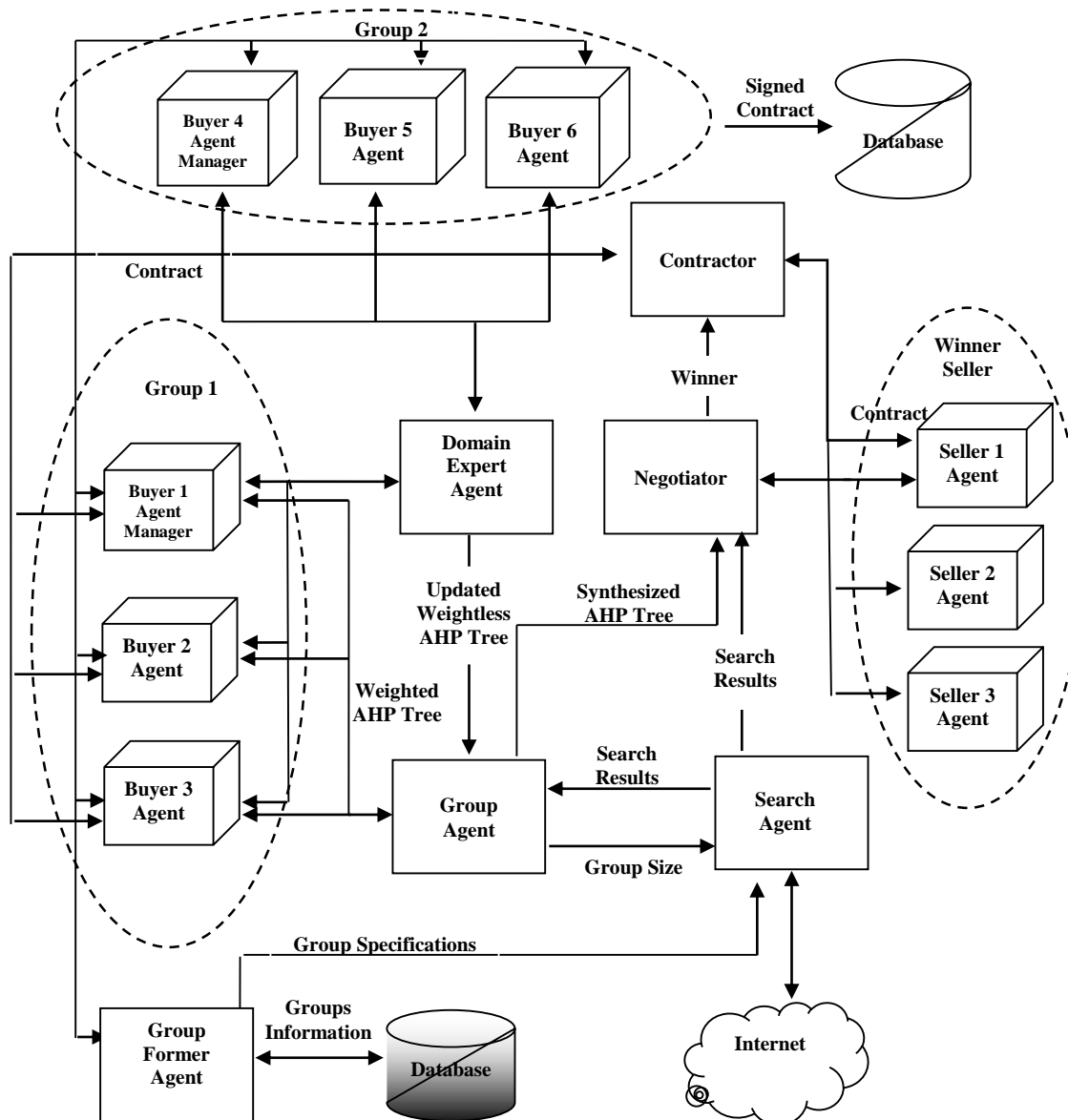


FIGURE 1: SYSTEM ARCHITECTURE

2.1. Constructing

In order to establish criteria for evaluating different products from different sellers, we need to determine *descriptive parameters* for the products of the same type, which in return will form AHP tree for scoring different products from different vendors. As also mentioned before, this can be performed either by an automated agent, or by domain experts who recognize what criteria may be important in assessing the value of a specified product such as a laptop. For example, they may recognize that Delivery Date is one of the important factors for customers and should be included in AHP tree for evaluating and scoring different products of the same type. After recognizing evaluation parameters for the products of a specific type, now we are ready to form *Weightless AHP tree* for evaluating and scoring products. After completing this step, *Weightless AHP Tree* will be provided.

2.2. Forming

During this step, purchasing groups should be formed. Now such questions may be made in your mind, so it's better to answer them before starting: what is a purchasing group? And why do we need to form it? Now we give a definition to a purchasing group in answer to the first question. A **Purchase Group** is a group of customers all agreed upon a range of products satisfying their requirements and expectations. A very important notation here is that, a Purchase Group will satisfy all customers' needs. In other words, when a customer agrees to join a specified group, he/she expresses that all products in that range exactly satisfy his/her needs, and there is no scarification of preferences. For the second questions, Forming a purchasing group leads to increasing bargaining power of customers and taking more discounts from potential sellers. Therefore, larger group sizes often cause better deals and more discounts for the members of that group. The user may decide either to create a new group or to join an existing group. Creating a new group occurs when the user has visited all groups and couldn't find a group which satisfies his/her needs. So he/she decides to create his/her own group and may invite the others (such as his/her friends, family and etc.) to join his group to form a larger group and to enhance the bargaining power of the group which leads to make a better deal. Notice that we have devised a **Recommender System** [19]. In our system which can find a group with most similarity to preferences of a user and help him/her to join the group which best satisfies his/her needs. Each group is managed by the initiator of that group whom has originally created it.

Having created the group, its manager is the only member. Now that he/she has created the group, he/she must go to Group Specification part to determine his/her preferences, which will act as the group specifications (see fig. 2).

FIGURE 2. PREFERENCES GATHERING SNAPSHOT

This page serves as a tool by which the manager specifies what exactly he/she wishes to buy, which also serves as the **group specifications**. In addition, as mentioned before there is no scarification of users' preferences because this system lets each user to express what exactly he/she wished to buy. As an example let's consider CPU as an evaluation criterion (CPU is one of Descriptive Parameters for a laptop just like Delivery Time). The user chooses a couple of CPU models from a list of CPU models automatically suggested (and also ranked) by the system. By choosing these CPU models, he/she states that his/her purchased laptop must incorporate one of these CPU models. *This is the real liberty in determining preferences offered by UUT-Trade*. After creating the group and specifying its overall specifications, this group will be obvious in the list of existing groups. Then the other users may visit group specifications of this group, by selecting it from the list of existing groups, and if they want, they may decide to join this group. Accordingly, either this group or other existing groups may be enhanced in members and become larger and larger or new groups may be created. Now, we are going to describe message structure of UUT-Trade, by presenting the Sequence diagrams, which describes messages exchanged between existing agents. For the sake of simplicity, the sequence diagram of the system will also be presented in a step by step manner. Message Structure of this step is presented in fig. 3.

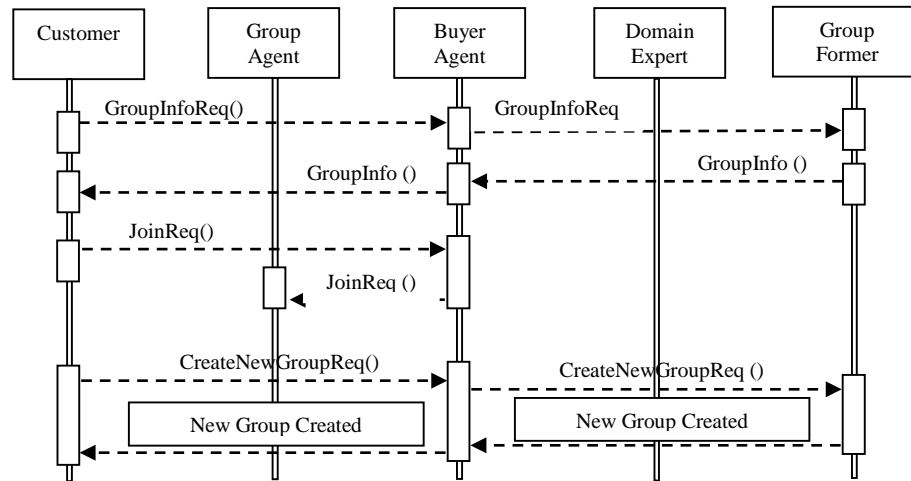


FIGURE 3: SEQUENCE DIAGRAM-FORMING

2.3. Updating

After Weightless AHP Tree has been formed, a specified automated agent (called Domain Expert Agent in the proposed system) may be assigned to solicit additional evaluation parameters from all users of a group to be added on Weightless AHP. For this sake, we can devise an “Others” branch on the level one of our AHP which will contain sub-branches identified by different users, to reflect updated requirements. This concept is demonstrated in fig. 4.

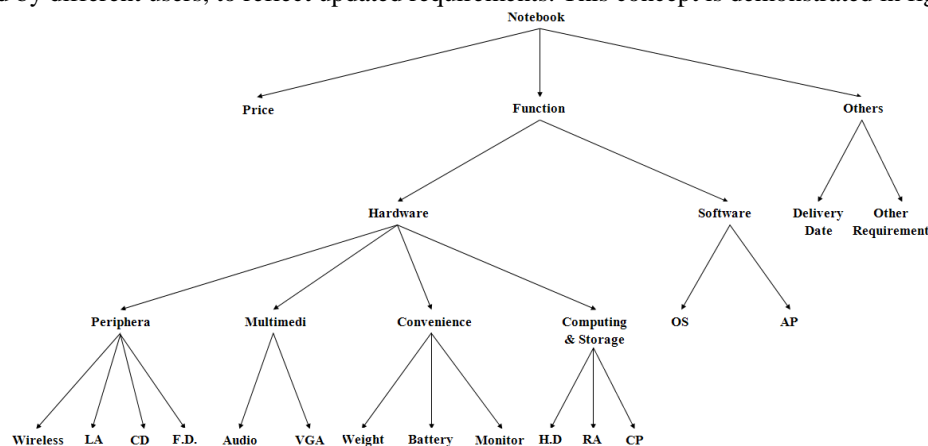


FIGURE 4: UPDATED AHP TREE

Message Structure of this step is presented in fig. 5.

2.4. Synthesizing

Now, we have got formed Weightless AHP Tree and (if required) updated it, first we should weight it and then synthesize it, before using it for scoring different products. We exploit an approach for weighting this tree, which can be done using a specific agent. This agent can gather customer’s opinions about each of evaluation criteria by *pair-wise comparisons* on each criterion and then synthesize them to obtain a unified score for that specified criterion. How this task can be done, and fundamental concepts of AHP, is explained in [6]. However, it’s better to cite the formula used to synthesize preferences. Assume that, $x^1_{ij}, x^2_{ij}, \dots, x^m_{ij}$ are m group members’ individual judgments on comparing criterion, i and j , then the aggregated judgment for that comparison is $(x^1_{ij} * x^2_{ij} * \dots * x^m_{ij})^{1/m}$ ([6]).

At the end of this step, we are provided with an AHP tree formed based on the criteria identified by *domain specialists* or automated agents, and weighted by kind of the opinions of all customers. Figure 6, shows Sequence diagram-synthesizing. This is made partially possible by synthesizing approach used in AHP method. However, we emphasize that this tree acts as the average of all customers’ opinions.

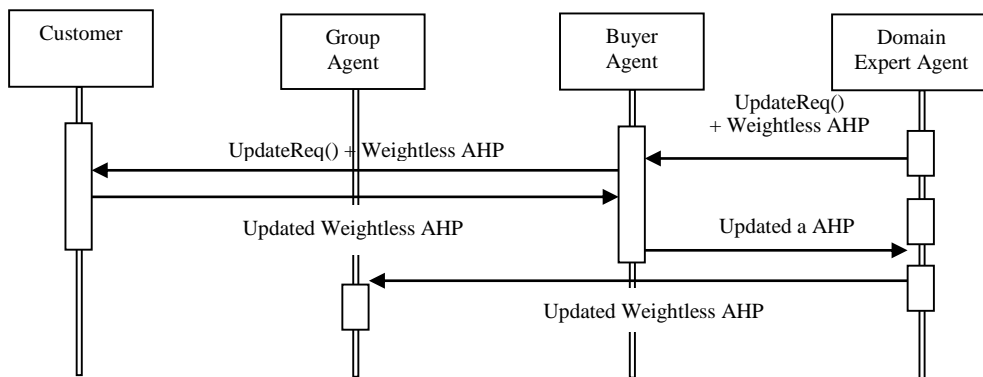


FIGURE 5: SEQUENCE DIAGRAM-UPDATING

The major output of this step is the **Synthesized AHP tree** for the specific product to be purchased. Message Structure of this step is presented in fig. 6.

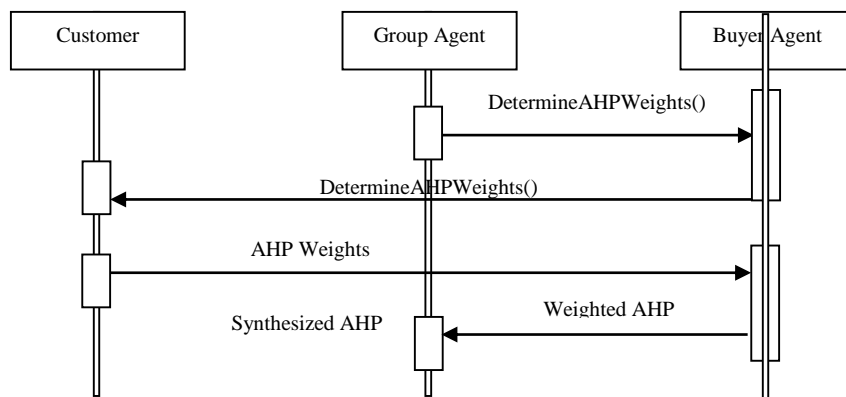


FIGURE 6: SEQUENCE DIAGRAM-SYNTHESIZING

2.5. Searching

After forming purchasing groups and determining their specifications by the group manager, a search agent should be responsible for searching on the internet and finding all existing products that match the group specifications. After doing this task, the system is provided with a number of potential sellers who carry the products which exactly fulfill the preferences of each group. Message Structure of this step is presented in fig. 7.

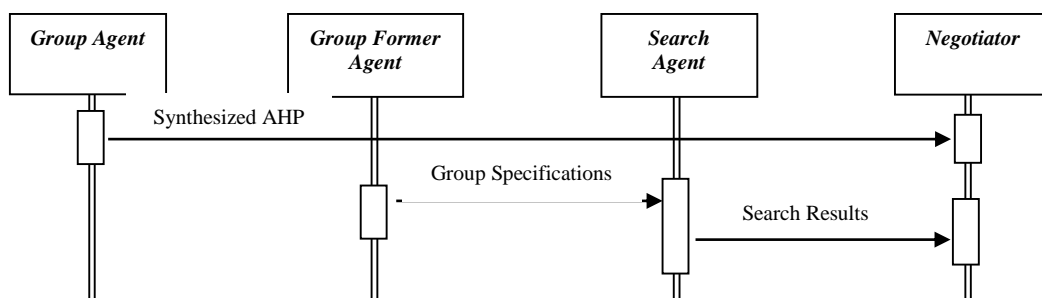


FIG. 7: SEQUENCE DIAGRAM-SEARCHING

2.6. Communicating

After the search agent finishes its task of finding matching products, and the search results are provided, the system should communicate the search results to individual group members and gain their approval. In other words, the users of a group should commit that they will accept the negotiation results. This commitment serves as a force on the users whom have signed it and makes them to buy the negotiated and selected product and prevents rejecting the selected product after negotiation has been done. This can be applied by making them to sign a **legal agreement**, such as the agreements some Internet sites present to their members asking them to respect the rules of the site. Message Structure of this step is presented in fig. 8.

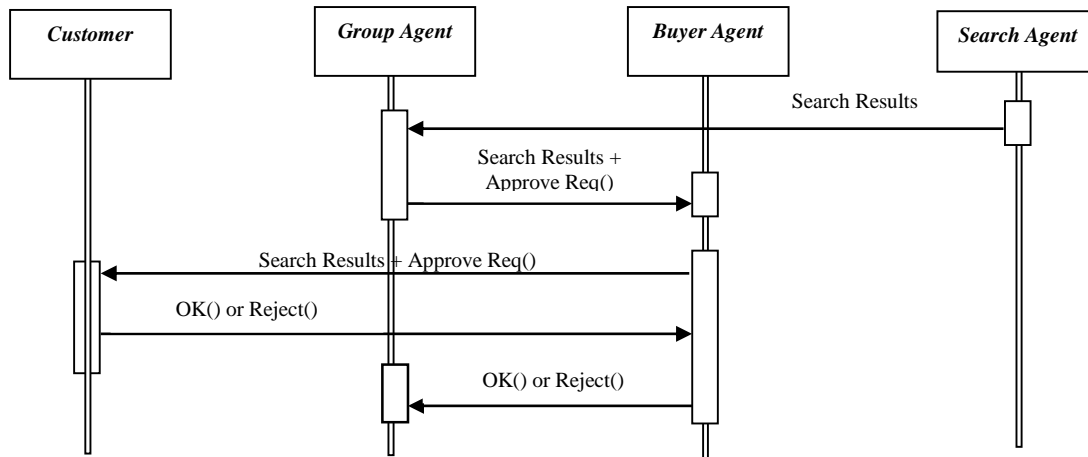


FIGURE 8: SEQUENCE DIAGRAM-COMMUNICATING

2.7. Calculating

By using the *Synthesized AHP tree* obtained before, now we can score each product matching to the group specifications. One specific agent should be assigned for calculating the score of each and then bargaining (which is the responsibility of *Negotiator Agent* in UUT-Trade). Calculating method to compute Final score of a product is the same as the one presented in [6]. This score is called *descriptive parameters score (DPS)*, which describes its quality in terms of descriptive parameters. Each merchant offers a price to its product, and this price can't be used to evaluate that product on its own because quality is also important. So we need the ratio of *Price to Descriptive Parameters Score (Pr/DPS)* to evaluate each product. This ratio in return will be used for bargaining.

2.8. Negotiating

This step is the most important step of UUT-Trade process. During this step, the system must negotiate with potential sellers and bargain with them. Bargaining will be accomplished based on the *Pr/DPS ratios* calculated as explained in section 2.7. Bargaining Algorithm is presented by the pseudo code, in fig. 9.

Bargaining algorithm	
00	int ratio[n], price[n], desc[n];
01	Bool selected[n] = {false}; //ratio[i]=price[i]/desc[i]
02	Sort desc[] in descending order;
03	Rearrange ratio[] & price[] based on the sorted desc[];
04	For i = 0 to i <= n
05	{
06	For t = 0 to t <= n
07	selected[t] = false;
08	while(true)
09	{
10	find j such that ratio[j] - ratio[i] is minimum &
11	ratio[j] < ratio[i] & selected[j] == false;
12	announce to i th seller that j th seller has offered a
13	better deal compared to him/her and encourage him/her to lower
14	his/her price to outdo j th seller in competition;
15	wait (a specific amount of time);
16	if(there is a bid from i th seller)
17	{
18	update(price[i]);
19	update(ratio[i]);
20	} else if(there is not another bid from i th seller)
21	selected[j] = true;
22	if(ratio[i] <= ratio[j])
23	selected[j] = true;
24	if(there is not another j with selected[j] == false)
25	break;
	}
	}
	}

FIGURE 9 : BARGAINING ALGORITHM PSEUDO CODE (C++)

This pseudo code explains how bargaining with potential sellers is implemented. This algorithm diminishes all prices as much as possible. desc[] and price[] and ratio[] are arrays storing “*descriptive parameters score(DPS)*” and “*Price*” and “*Pr/DPS ratio*” respectively. Message Structure of this step is presented in fig. 10.

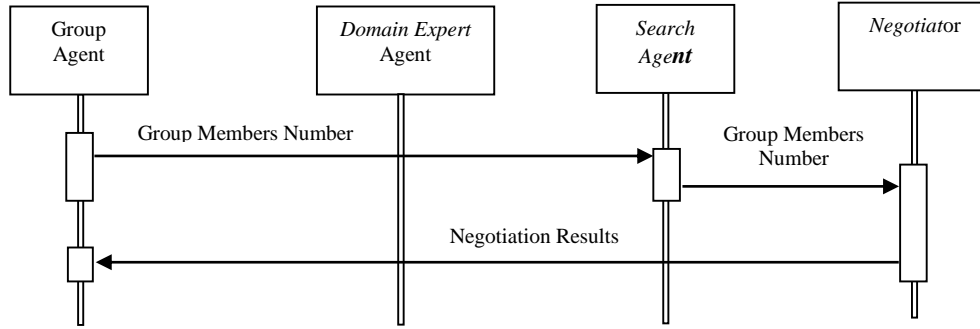


FIGURE 10: SEQUENCE DIAGRAM-NEGOTIATING

2.9. Voting

After completing bargaining, each group applies a voting procedure to determine which potential product will be bought. Before starting to present the voting formula, we are required to define some variables. Let P_{ij} , be the percentage, given by i^{th} user to j^{th} product, which describes how much i^{th} user is interested in buying j^{th} product in terms of percent. Furthermore, assume SC_j , as the final score for j^{th} product. We define the variable W as below:

$$W = \sum_{j=1}^m SC_j \quad (1)$$

Voting will be performed weightily, and the weights will be determined based on the proportion of variable SC_j to variable W . In order to accomplish voting, each user will be presented with a form which solicits his/her interest to buy a product, in terms of percent. Let n be the number of users voting on m products. We use a formula to calculate a number which will finally present the score of product in the ranking. For each product, we should calculate a value based on the formula presented below:

$$\forall j : 1 \leq j \leq m : S_j = \left(\frac{(\sum_{i=1}^n P_{ij})}{n} \right) * \frac{SC_j}{W} \quad (2)$$

- P_{ij} , is the percentage, given by i^{th} user to j^{th} product, in terms of percent.
- SC_j , is the final score for j^{th} product, calculated based on the Synthesized AHP in section 2.2.7. We called it *Descriptive Parameters Score (DPS)*, before.
- S_j , is the final score of j^{th} product, determining the final product to be bought.
- m , is the number of products.
- n , is the number of users.

After calculating S_j for each j **between 1 and m**, the product having the greatest S_j , should be selected and will eventually be bought. Message Structure of this step is presented in fig. 11.

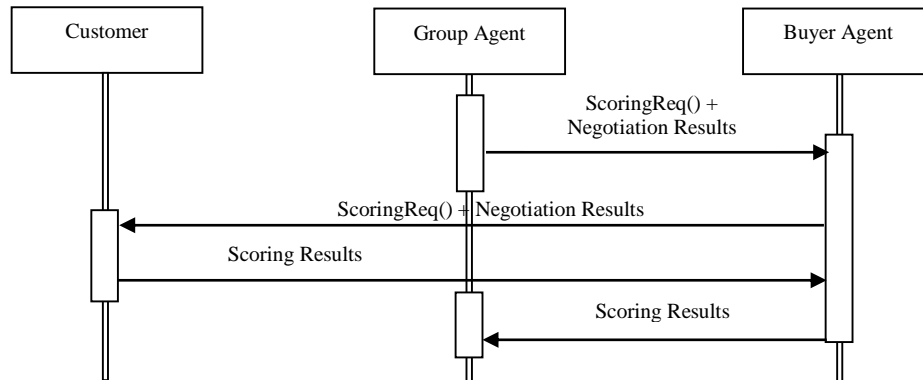


FIGURE 11: SEQUENCE DIAGRAM-VOTING

2.10. Contracting

During this step, a contract must be provided by the contractor agent, with the winner for each purchasing. Contract will be signed by the winner and also individual group members. After that, payments in addition will be settled and hereafter, the products are ready to be delivered by the seller. Message Structure of this step is presented in fig. 12.

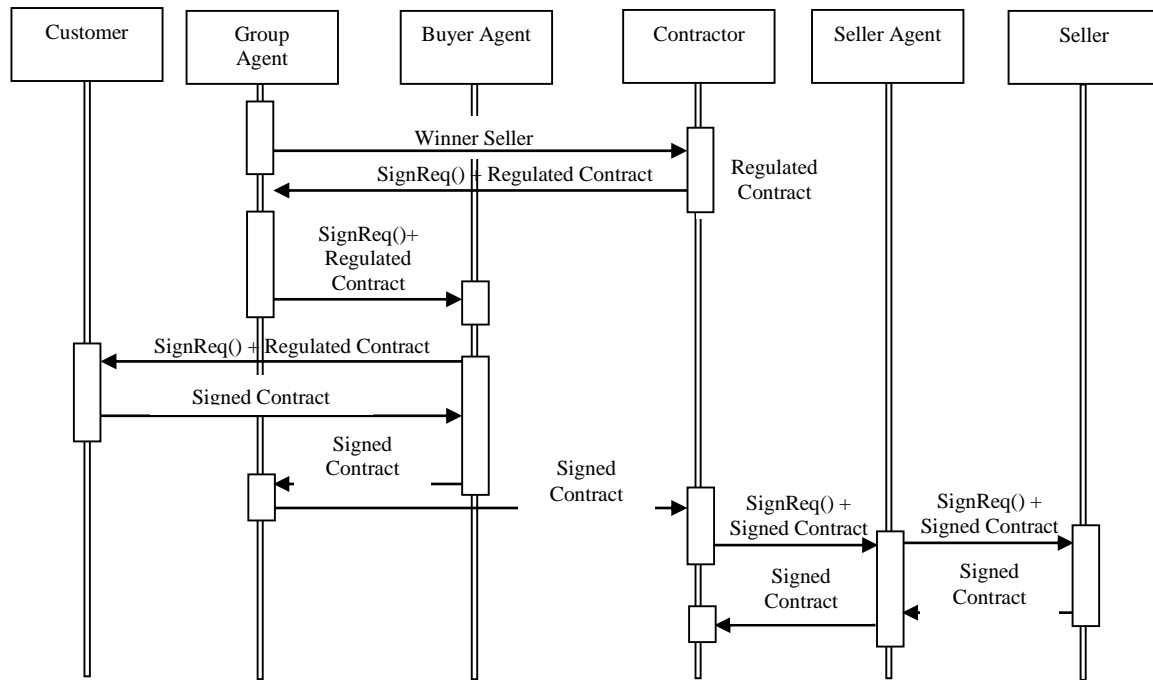


FIGURE 12: SEQUENCE DIAGRAM-CONTRACTING

IV. EVALUATION

In our experiment we compared proposed model with two models: BCP model proposed by [6] and Negotiation system presented in [20] which are described in section 2. We evaluated performance of the proposed negotiation system based on user satisfaction. The user satisfaction effects on performance of system. Performance is affected on system success because users can achieve high results by using the system [6]. We use online questionnaire for assessment of this system. The questions are about sub factors of user satisfaction (Ease of use, Timeliness and Accuracy) these factors are described as below:

Ease of use is the degree to which the person believes that using the system is free of effort.

Timeliness is degree to which the system provides on-time and up to date information to the customers.

Accuracy is the degree to which the outcome of the negotiation matches user's needs.

A total of 214 user participated in the experiment, but only 127 responses are usable to evaluate the system. We rated to each question a score between 4 to 10 then we calculate average rate of each factor. Some of questions of questionnaire are reflected in Table 1.

TABLE 1. DESCRIPTION OF CONSTRUCTS AND ITEMS

Measures	Item Code	Original Instrument Item
Accuracy	A1	Do the requirements accurately reflect the wishes and needs of the stakeholders?
	A2	Did the outcome of the negotiation match what you thought it would be before you began exchanging offers?
Ease of use	E1	Is the system user friendly?
	E2	Have all relevant requirements for the system to be developed been documented?
Timeliness	T1	Do you get the information you need in time?
	T2	Does the system provide up to date information?

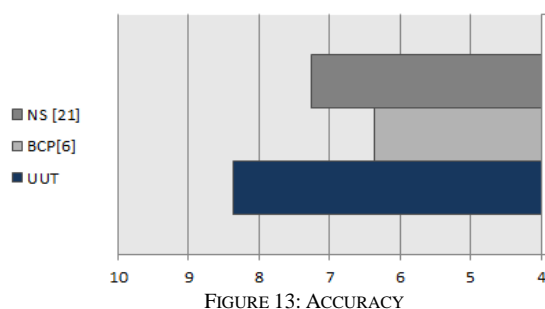


FIGURE 13: ACCURACY

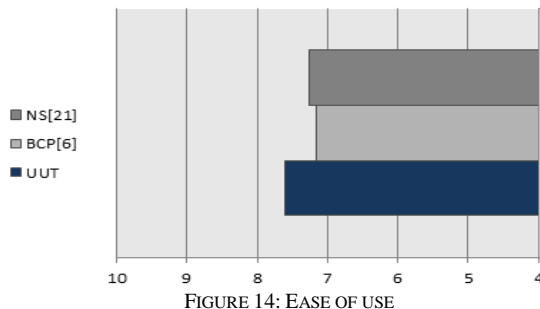


FIGURE 14: EASE OF USE

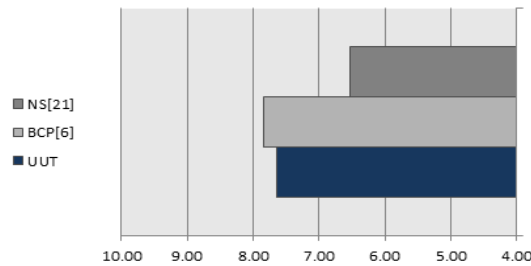


FIGURE 15: TIMELINESS

The results show that accuracy of UUT system is more than BCP [6] and Negotiation System in [20] as shown in Figure 13 and the Figure 15 shows that timeliness of UUT system is less than two Other systems but ease of use of proposed system is higher than two other systems (Fig. 14). This means that the user satisfaction in proposed model is enhanced compared to two other models. User satisfaction depends on some variable that we use Ease of use, accuracy and timeliness of them. In proposed system Ease of use and accuracy are increased but the timeliness is decreased slightly than two other systems. Our goal is to increase the accuracy of the proposed system in which it is grown, but it is caused that timeliness of proposed system reduced. The average score of accuracy in proposed system is 8.377 while this score in the BCP model is 7.267 and the Negotiation System in [20] is 6.362. Accuracy is one of the important factors to user satisfaction.

V. CONCLUSION

In this paper, we proposed a flexible system called UUT-Trade, which satisfies all users' preferences and suggests no scarification of users' needs. By this system, we explained how we improved BCP model proposed by [6], in our system, and how we used this model to enhance the bargaining power of customers, while we eliminated the drawback of scarifying users' preferences from their proposed system. We used AHP tree in a different way, to evaluate the quality of a product. That is, we used AHP tree to obtain descriptive scores (DPS) for each of the negotiated products, unlike the authors [6] whom used AHP tree to synthesize users' preferences which was the main reason of the scarification occurrence. Besides, we eliminated the drawback of their negotiation mechanism, which was limited to only negotiating with manufacturers. Using our proposed negotiation mechanism, you will be able to negotiate with anybody, addressing the preferences of purchasing groups (such as manufacturers, retailers, etc). UUT-Trade system used a new negotiation algorithm, which diminishes all prices, and then the users have got liberty to choose between potential sellers by performing a weighted voting. In the future, we hope, we can propose the particular approach for payments of our system. Besides, the Recommender System, devised in our system, is also worthwhile to be investigated and extended, to achieve better performance and further enhancements in the users' satisfaction.

VI. REFERENCE

- [1] Guo, X., & Lim, J. (2012). Decision support for online group negotiation: Design, implementation, and efficacy. *Decision Support Systems*, 54(1), 362-371.
- [2] Uzoka, F. M. (2005). AHP-based system for strategic evaluation of financial information. *Information-Knowledge-Systems Management*, 5(1), 49-61.
- [3] Leung, P., Muraoka, J., Nakamoto, S. T., & Pooley, S. (1998). Evaluating fisheries management options in Hawaii using analytic hierarchy process (AHP). *Fisheries Research*, 36(2), 171-183.
- [4] Saaty, T. L. (1994). Highlights and critical points in the theory and application of the analytic hierarchy process. *European journal of operational research*, 74(3), 426-447.
- [5] Schmitt, C., Dengler, D., & Bauer, M. (2002, October). The maut-machine: an adaptive recommender system. In *Proceedings of ABIS Workshop*, Hannover, Germany.
- [6] Chen, D. N., Jeng, B., Lee, W. P., & Chuang, C. H. (2008). An agent-based model for consumer-to-business electronic commerce. *Expert Systems with Applications*, 34(1), 469-481.
- [7] Cheung, S. O., Yiu, K. T., & Suen, H. (2004). Construction negotiation online. *Journal of construction engineering and management*, 130(6), 844-852.
- [8] de Campos, L. M., Fernández-Luna, J. M., & Huete, J. F. (2008). A collaborative recommender system based on probabilistic inference from fuzzy observations. *Fuzzy Sets and Systems*, 159(12), 1554-1576.
- [9] Zhen, L., Huang, G. Q., & Jiang, Z. (2009). Recommender system based on workflow. *Decision Support Systems*, 48(1), 237-245.
- [10] Hammer W.C, Clay. G.A (1969), *The Economics of Bargaining*, University of Georgia Press, Athens.
- [11] Raiffa, H. (1982). *The art and science of negotiation*. Harvard University Press.
- [12] Huang, C. C., Liang, W. Y., Lai, Y. H., & Lin, Y. C. (2010). The agent-based negotiation process for B2C e-commerce. *Expert Systems with Applications*, 37(1), 348-359.
- [13] He, M., Jennings, N. R., & Leung, H. F. (2003). On agent-mediated electronic commerce. *Knowledge and Data Engineering, IEEE Transactions on*, 15(4), 985-1003.
- [14] Kersten, G. E., & Lai, H. (2007). Negotiation support and e-negotiation systems: an overview. *Group Decision and Negotiation*, 16(6), 553-586.
- [15] Mustajoki, J., & Hamalainen, R. P. (2000). Web-HIPRE: Global decision support by value tree and AHP analysis. *INFOR J*, 38(3), 208-220.
- [16] <http://hipre.aalto.fi/>
- [17] Schleiffer, R. (2005). An intelligent agent model. *European Journal of Operational Research*, 166(3), 666-693.
- [18] Krovi, R., Graesser, A. C., & Pracht, W. E. (1999). Agent behaviors in virtual negotiation environments. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 29(1), 15-25.
- [19] Choi, S. P., Liu, J., & Chan, S. P. (2001). A genetic agent-based negotiation system. *Computer Networks*, 37(2), 195-204.
- [20] Lee, W. P. (2004). Towards agent-based decision making in the electronic marketplace: interactive recommendation and automated negotiation. *Expert Systems with Applications*, 27(4), 665-679.

- [21] Luo, X., Jennings, N. R., Shadbolt, N., Leung, H. F., & Lee, J. H. M. (2003). A fuzzy constraint based model for bilateral, multi-issue negotiations in semi-competitive environments. *Artificial Intelligence*, 148(1), 53-102.
- [22] Maes, P., Guttman, R., & Moukas, A. G. (1998). Agents that buy and sell: Transforming commerce as we know it. *Communications of the ACM*.
- [23] Lau, R. Y. (2007). Towards a web services and intelligent agents-based negotiation system for B2B eCommerce. *Electronic Commerce Research and Applications*, 6(3), 260-273.
- [24] Al-Sakran, H. B2C E-Commerce Fact-Based Negotiation Using Big Data Analytics and Agent-Based Technologies.
- [25] Haghighatjoo, M., Masoumi, B., & Meybodi, M. R. (2014, October). Strategic agents for multi resource negotiation using learning automata and case-based reasoning. In *Computer and Knowledge Engineering (ICCKE), 2014 4th International eConference on* (pp. 35-41). IEEE.
- [26] Kasperczyk, N., & Knickel, K. (2004). "Analytic hierarchy process (AHP)," IVM Institute, [Online]. Available: http://www.ivm.vu.nl/en/Images/MCA3_tcm53-161529.pdf.
- [27] Omar, M. K. (2011, September). The solutions of DT and AHP for supplier selection problem. In *Quality and Reliability (ICQR), 2011 IEEE International Conference on* (pp. 506-510). IEEE.

Parallelizing K-Way Merging

Hazem M Bahig^{1,2} and Ahmed Y Khedr^{1,3}

¹College of Computer Science and Engineering, Hail University, Hail, KSA.

²Department of Mathematics, Faculty of Science, Ain Shams University, Cairo, Egypt.

³Systems and Computer Department, Faculty of Engineering, Al-Azhar University, Cairo, Egypt.

Abstract— The k -way merging problem is to find a new sorted array as an output from k sorted arrays as an input. In this paper, we consider the elements of the k sorted arrays are data record, where the value of the key for each record is a serial number. The problem is used to design efficient external sorting algorithm. We proposed two optimal parallel algorithms for k merging. The first one is based on merging k sorted arrays of n records in a new sorted array of length n . The second one is based on merging k sorted arrays of n records in a new sorted array of length $n+o(n)$ which is called padded merging. The running time for each algorithm is $O(\log n)$ and $O(1)$ under EREW and CRCW PRAM respectively.

Keywords- merging; k -merging; padded merging; PRAM; optimal algorithm; parallel algorithm.

I. INTRODUCTION

Given k sorted arrays of total size n as an input. The k -way merging problem is to produce a single new sorted array, A , contains all the elements of the input. In case of $k=2$, the problem is called the binary merging problem or merging problem.

In general, the merge problem plays an important step in solving many applications in the field of computer science such as sort, reconstruction of large phylogenetic trees, and database management systems [5][13][16][18]. One of these important applications is the merge sort algorithm. In merge sort algorithm, we divide the original array into two equal size subarrays and then sort each subarray recursively. After that, we merge the two sorted subarrays.

In many applications, the data to be sorted is too large and therefore we cannot fit the data in the internal memory. In this case, the data will be store in the external storage, such as a hard disk. But the performance of the optimal merge-sort algorithm is not well in case of the data stored in external storage. Because the reading and writing from and to the external storage is very slow. In this case, the k -way merging algorithm is an efficient technique to sort the data in the external storage and the sorting problem is called external sorting.

The merging problem has been studied by many researchers on sequential and parallel platforms. The summary of these researches is given in Table 1. In this summary, we focused only on the shared memory model especially parallel random access machine. In the table we use p to represents the number of processors. We also use two terms, work and cost. The work of the algorithm is the total number of operations done by all processors, while the cost of the algorithm is the product of running time and the number of processors. Also we use $\alpha(n)$ to represents the inverse of Ackermann's function.

From the table we observe the following for the algorithm under PRAM.

1. We can merge two sorted array in constant time in some special cases and $p=n$ as in[3][4].
2. The optimal merging algorithm without any restrictions on the input has running time $O(\log n)$ and $O(\log \log n)$ under EREW and CREW respectively.
3. The optimal work merging algorithm for integer numbers has running time $O(\log \log n + \log \text{Min}\{n, m\})$ and $O(\alpha(n))$ under EREW and CREW respectively, where m is the domain of integer, $[1, m]$.
4. The optimal work k merging algorithm has running time $\Omega(\log n)$ and $\Omega(\log \log n + \log k)$ under EREW and CREW respectively.

In this paper, we study the k merging problem on PRAM. In some applications, such as external sorting, the elements of the k sorted arrays are records and the records are sorted according to the primary key. We proposed two k merging algorithms under EREW and CRCW PRAM. The first algorithm merges the k sorted arrays of size n in a new sorted array of size n . The second

algorithm merges the k sorted arrays of size n in a new array of size $n+o(n)$. In case of EREW PRAM, the algorithm runs in logarithmic time, while the algorithm runs in constant time in case of CRCW PRAM.

Table 1: Comparison between merging algorithms

Ref.	Input	p	Model	Time	Work	Cost	Comments
[12]	2 sorted arrays	1	Sequential	$O(n)$	$O(n)$	$O(n)$	---
[10]	k sorted arrays	$n/\log n$	EREW PRAM	$O(\log n)$	$O(n)$	$O(n)$	---
[14]	2 sorted arrays	$n/\log\log n$	CREW PRAM	$O(\log\log n)$	$O(n)$	$O(n)$	---
[9]	2 sorted arrays	n	EREW PRAM	$O(\log\log n + \log \min\{n, m\})$	$O(n)$	$O(n \log\log n)$	Integers
[3][4]	2 sorted arrays	p	EREW PRAM	$O(n/p)$	$O(n)$	$O(n)$	Special case
[6]	2 sorted arrays	$n/\log\log\log m$	CREW PRAM	$O(\log\log\log n)$	$O(n)$	$O(n)$	Integers
[6]	2 sorted arrays	$n/\alpha(n)$	CREW PRAM	$O(\alpha(n))$	$O(n)$	$O(n)$	Integers
[10]	k sorted arrays	$n/\log n$	EREW PRAM	$O(\log n \log k)$	$O(n \log k)$	$O(n \log k)$	Integers
[17]	k sorted arrays	$(n \log k)/\log n$	CREW PRAM	$O(\log n)$	$O(n \log k)$	$O(n \log k)$	Integers
[11]	k sorted arrays	n	EREW PRAM	$\Omega(\log n)$	$O(n \log k)$	$O(n \log n)$	---
[11]	k sorted arrays	n	CREW PRAM	$\Omega(\log\log n + \log k)$	$O(n \log k)$	$O(n \log n)$	---

The research paper consists of an introduction and four sections. In Section II, we give the definition of problem and the model of computation used. In Section III, we describe the main idea, steps, and the complexity analysis of the proposed algorithm under EREW and CREW PRAM. In Section IV, we extend the domain of the primary key and then modified the algorithm. Finally in Section V, we show the conclusion of our work.

II. PRIMELIARY

In this section, we give a brief description about the parallel model used in designing the algorithm and the complete description of our problem.

A. Parallel Random Access Machine

A Parallel Random Access Machine, PRAM, is the natural extension of the universal model of sequential machine Random Access Machine, RAM. Also, the model is a type of shared memory Single Instruct Multi Instruction, SIMD. It consists of p identical RAM processors and large M shared memory cells. The p processors operate synchronously and communicate through the shared memory. Each processor p_i may execute (i) read from a shared memory cell, (ii) write to a shared memory cell, and (iii) local computation.

Due to the memory access conflicts in shared memory for reading and writing, three realistic mechanisms are proposed.

- Exclusive Read Exclusive Write (EREW) PRAM: no simultaneous read or write by two or more processors from or to the same memory cell location.
- Concurrent Read Exclusive Write (CREW) PRAM: simultaneous reads of the same memory cell by two or more processors allowed, but no simultaneous writes by two or more processors to the same memory cell location.
- Concurrent Read Concurrent Write (CRCW) PRAM: simultaneous reads or writes from or to the same memory cell by two or more processors allowed.

In CRCW, different submodels are proposed to illustrate the mechanisms of CW. In our proposed algorithm we use the Common CRCW. In a Common CRCW PRAM, concurrent writes are allowed only if all processors have the same value at the same time.

B. Problem Formulation and Related Problem

We can formulate the problem of k merging records of serial numbers as follows.

Given k sorted arrays of data records, $R_i = (r_{i0}, r_{i1}, \dots, r_{i(n_i-1)})$, $0 \leq i < k$ such that: (1) the elements of the array R_i is sorted based on the field "key". I.e $r_{ij} \cdot \text{key} < r_{i(j+1)} \cdot \text{key}$, $0 \leq j < n_i - 1$ and $0 \leq i < k - 1$. (2) The values of keys in all k arrays are serial number. (3) The total number of records is $n = n_1 + n_2 + \dots + n_k$. The output of the k -way merging is a new sorted array of records $R = (r_0, r_1, \dots, r_{n-1})$ such that $r_i \cdot \text{key} < r_{i+1} \cdot \text{key}$, $0 \leq i < n - 1$.

In our proposal algorithm, we need the problem of finding the minimal and maximal elements, so the optimal results for this problem on different models of PRAM are as follows,

Proposition 1 [2]: The problem of computing the maximum/minimum of n elements in an array A can be performed in $O(n/p)$ time using p EREW PRAM processors, for $p \leq n/\log n$.

Proposition 2 [2]: The problem of computing the maximum/minimum of n elements in an array A can be performed in

- $\tilde{O}(1)$ time using n CRCW PRAM processors.
- $O(\log \log n)$ time using $n/\log \log n$ Common CREW PRAM processors.

Proposition 3 [2]: The maximum/minimum of n integers in the range $[1, n^{O(1)}]$ can be found in $O(1)$ time using n CRCW PRAM processors.

III. OPTIMAL K-WAY PARALLEL MERGING ALGORITHM

In this section we present the main idea and the steps to design an optimal parallel algorithm to merge k sorted arrays. We also analyze the algorithm based on two different models of PRAM: EREW and CRCW.

A. Main Idea

Since the elements of each array are sorted based on the field *key*, and the values of the keys are serial number. We can map the values of the keys of the records into an integer range. So, each record can be representing as an integer number. We can do this process by applying a mapping function that maps the n records into the domain $[0, n-1]$. After that we have n elements of consecutive integer numbers. Therefore, we can apply the address strategy to fill the record of key correspond to value i in the address i in the output array. The address or index strategy is used in many previous algorithms such as count sort and bit-index sort [19]. Figure 1 represents the idea of the proposed algorithm. In the figure, we have three sorted arrays, R_0 , R_1 , and R_2 , of lengths 6, 4, and 5 respectively. Each element in the arrays consists of two fields. The first one is the key of the record, while the other is the reminder data of the record.

B. Steps of k -way Merging Parallel Algorithm

We give here the main steps to merge k sorted arrays of records such that the keys of the records are serial numbers and consecutive values. The algorithm consists of three main steps as follows.

Step 1: Determine the minimum, \min , value for all the keys in the k sorted arrays in parallel.

Step 2: Compute the address array AR_i for each sorted array R_i in parallel as follows:

$$ar_{ij} = M(r_{ij} \cdot \text{key})$$

where M is the mapping function and defined as :

$$M(r_{ij} \cdot \text{key}) = r_{ij} \cdot \text{key} - \min$$

For all $0 \leq i < k$ and $0 \leq j < n_i$.

Step 3: For each element r_{ij} in the sorted array R_i , insert r_{ij} in the correct position in the output array R as follows.

$$r(ar_{ij}) = r_{ij}$$

Remark 1: we can combine the two steps 2 and 3, in one step as follows.

$$r(M(r_{ij} \cdot \text{key})) = r_{ij}$$

C. Complexity Analysis

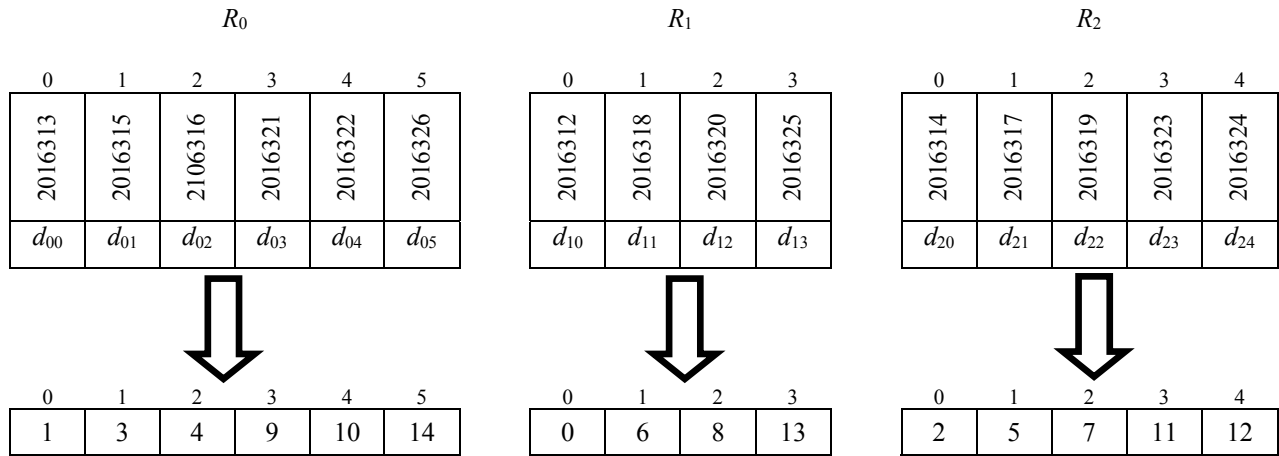
In this section we analyze the proposed algorithm according to time, number of processors, cost, and optimality. The analysis of the algorithm depends on the type of the model used.

In case of EREW and the number of processor $p=n/\log n$, the running time for Steps 1, 2, and 3 are $O(\log n)$, $O(1)$, $O(\log n)$ respectively. The overall running time for the proposed algorithm is $O(\log n)$. Therefore, the cost of k -way merging is $O(n)$ and the algorithm is optimal.

In case of CRCW and the number of processor $p=O(n)$, the running time for Steps 1, 2, and 3 are $O(1)$, $O(1)$, $O(1)$ respectively. The overall running time for the proposed algorithm is $O(1)$. Therefore, the cost of k -way merging is $O(n)$ and the algorithm is optimal.

Remark 2: if the minimum number is known, we can merge the k arrays in constant time under EREW PRAM.

Stage 1:



Stage 2:

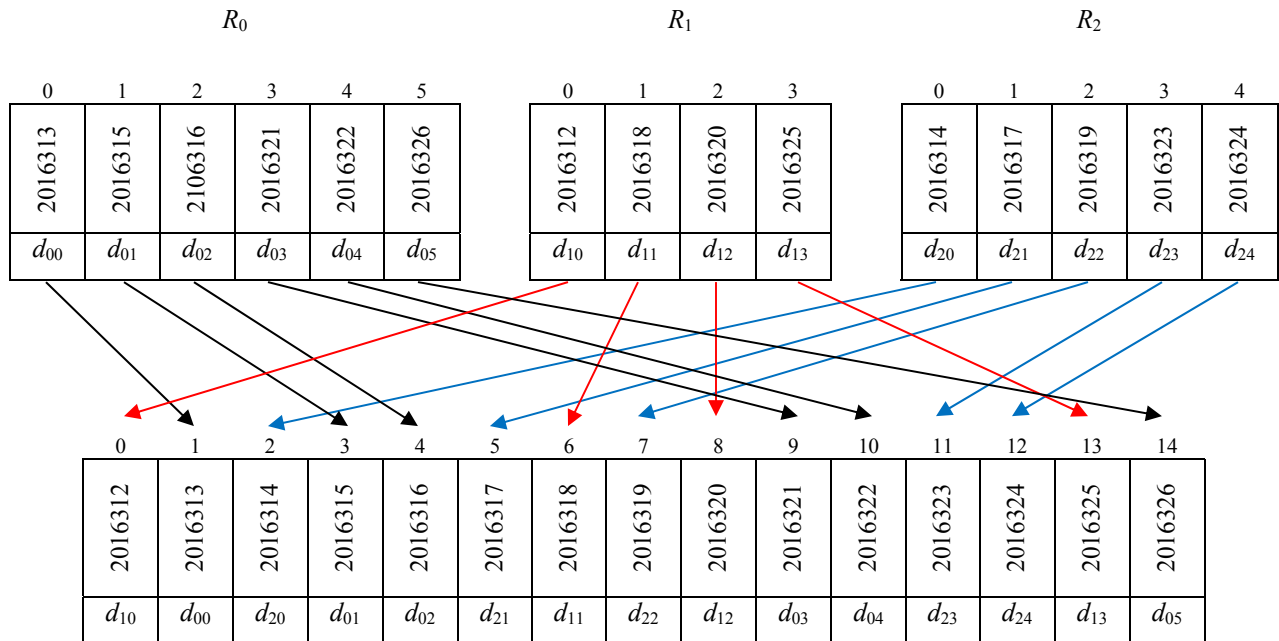


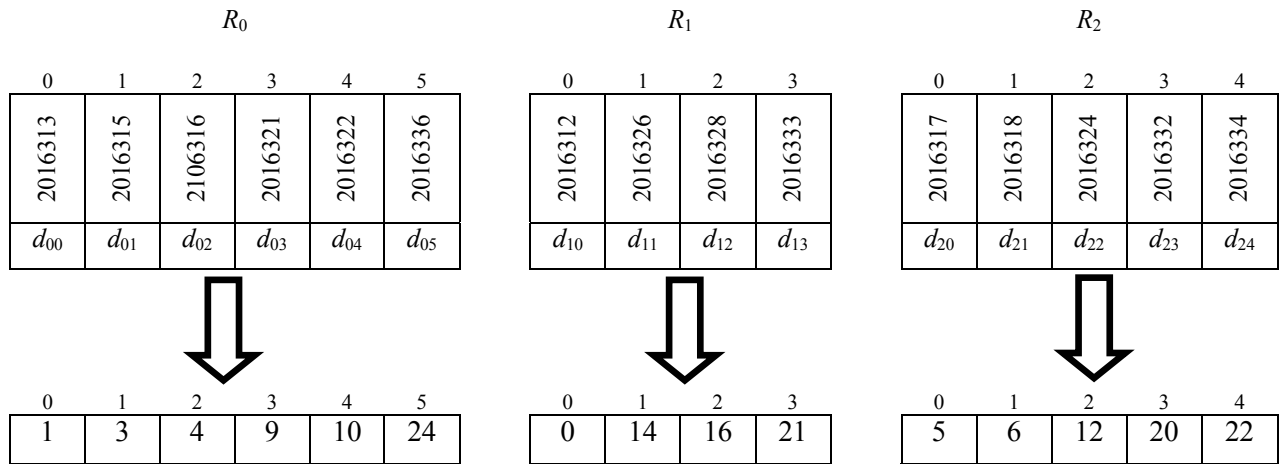
Figure 1: Two main stages for the proposed algorithm.

IV. OPTIMAL K-WAY PADDED PARALLEL MERGING ALGORITHM

In this section we study the same problem when the key of the records is not necessary consecutive. In this case, we use the same idea in previous section, but we use extra gap spaces in the output. Adding more extra gap spaces in the output is called padded technique. The concept of padded is used in different problems such as sorting [20][20]. In padded sort, we have n elements that are taken from a uniform distribution and we want to ordered the n values in array of length $n+o(n)$ such that all $o(n)$ locations are filled with NULL. So, we can apply this concept to our problem to merge the k sorted arrays in a new sorted array of length $n+o(n)$.

The padded concept is based on using extra space present the main idea and the steps to design an optimal parallel algorithm to merge k sorted arrays. We also analyze the algorithm based on two different models of PRAM: EREW and CRCW.

Stage 1:



Stage 2:

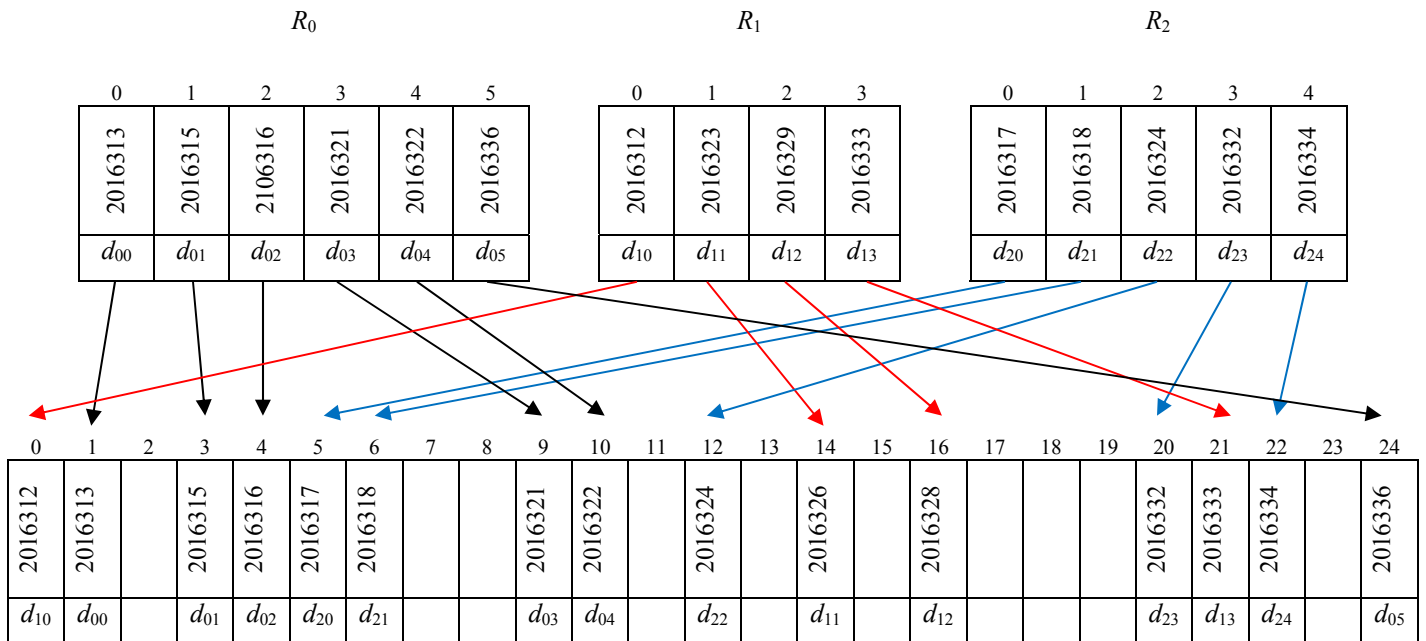


Figure 2: Two main stages for the proposed algorithm.

A. Main Idea

We use an address array of size m , where $m > n$. Because the values of the keys not necessary consecutive. We can map the values of the keys of the records into an integer range $[0, m-1]$. The value of m is equal to the difference between the largest and smallest values of the keys. Figure 2 represents the idea of the proposed algorithm, where the minimum value of the keys is 2016312 and the maximum value of the keys is 2016336.

B. Steps of k -way Padded MergingParallel Algorithm

The algorithm consists of three main steps as follows.

Step 1: Determine the minimum, min , and the maximum, max , values for all the keys in the k sorted arrays in parallel.

Step 2: Compute the address array AR_i for each sorted array R_i in parallel as follows:

$$ar_{ij} = M(r_{ij} \cdot \text{key})$$

where AR_i is array of length $max - min$.

Step 3: For each element r_{ij} in the sorted array R_i , insert r_{ij} in the correct position in the output array R as follows.

$$r(ar_{ij}) = r_{ij}$$

The proposed algorithm has the same running time as in the previous section under both models of PRAM.

V. CONCLUSION

In this paper we addressed the problem of merging k sorted arrays of records. Our study focused when the key of the record represents as a serial numbers. We proposed two algorithms under EREW and CRCW PRAM. The first one when the keys are consecutive, while the second when the keys are not necessary consecutive. The running time of the proposed algorithm is constant under CRCW PRAM, while the running time is $O(\log n)$ under EREW PRAM.

ACKNOWLEDGMENT

This research was supported by Research Deanship, Hail University, KSA, on grant R2-2013-CS-4.

REFERENCES

- [1] S. Akl. Parallel sorting algorithms. Academic Press, Orlando, 1985.
- [2] S. Akl. Parallel computation: models and methods. Prentice Hall, Upper Saddle River, 1997.
- [3] H Bahig. Parallel merging with restrictions. The Journal of Supercomputing, 43 (1): 99-104, 2008.
- [4] H Bahig. Integer merging on PRAM. Computing, 91(4), 365-378, 2011.
- [5] J. Bang-Jensen, J. Huang, and L. Ibarra. Recognizing and representing proper interval graphs in parallel using merging and sorting. Discrete Applied Mathematics 155(4):442-456, 2007.
- [6] O. Berkman, and U. Vishkin. On parallel integer merging. Information and Computation 106:266-285, 1993.
- [7] Th. Cormen, C. Leiserson, R. Rivest, and C. Stein. Introduction to algorithms. MIT, Cambridge, 1990.
- [8] E. Dekel and I. Ozsvath, Parallel external sorting. Journal of Parallel and Distributed Computing, vol. 6, 623-635, 1989.
- [9] T. Hagerup, and M. Kutylowski. Fast integer merging on the EREW PRAM. Algorithmica, 17:55-66, 1997.
- [10] T. Hagerup, and C. Rub. Optimal merging and sorting on the EREW PRAM. Information Processing Letters, 33:181-185, 1989.
- [11] T. Hayashi, K. Nakano, and S. Olariu. Work-time optimal k -merge algorithms on the PRAM. IEEE Transaction on Parallel and Distributed Systems, 9(3): 275-282, 1998.
- [12] R. Karp, and V. Ramachandran. Parallel algorithms for shared-memory machines. In: Van Leeuwen J (ed) Handbook of theoretical computer science, Vol A: Algorithms and complexity. Elsevier, Amsterdam, 869-941, 1990.
- [13] D. Knuth. The art of computer programming: sorting and searching. Addison-Wesley, Reading, 1973.
- [14] C. Kruskal. Searching, merging, and sorting in parallel computation. IEEE Transaction on Computers, 32(10):942-946, 1983.
- [15] T. Merrett. Relational information systems. Reston Publishing Co., Reston, 1984.
- [16] S. Olariu, C. Overstreet, and Z. Wen. Reconstructing binary trees in doubly logarithmic CREW time. Journal of Parallel and Distributed Computing, Vol. 27, 100-105, 1995.
- [17] Z. Wen. Multi-way merging in parallel. IEEE Trans. Parallel and Distributed Systems, vol. 7, no. 1, 11-17, Jan. 1996.
- [18] P. Valduriez, and G. Gardarin. Join and semijoin algorithms for multiprocessors database machines. ACM Transaction Database System 9:133-161, 1984.
- [19] L. F. Curi-Quintal, J. O. Cadenas, and G. M. Megson. Bit-index sort: A fast non-comparison integer sorting algorithm for permutations. International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAECE), 83 - 87, 2013.

- [20] P. D. MacKenzie and Q. F. Stout. Ultra-Fast Expected Time Parallel Algorithms. *Journal of Algorithms* 26 (1998), . 1–33.

Extended Smart Metering Display for Improved Energy Economy

Nisar Ahmed ¹, Muzafar Khan ², Muhammad Tahir ³, Shahid Yousaf ¹

¹ School of Engineering, Blekinge Institute of Technology, Karlskrona, Sweden

² College of Computer and Information Sciences (Muzahmiyah Branch), King Saud University, Riyadh, Saudi Arabia

³ Faculty of Computing and Information Technology, University of Jeddah, Jeddah, Saudi Arabia

Abstract Human dependency on technology is increasing day by day and environmental conditions are getting worse as a result. Energy consumption is increasing while the traditionally available energy sources like oil and gases are depleting. One of the major consumers is the domestic consumer, who plays the least part in energy management. One way to increase efficiency in energy management is, therefore, to pass part of it to the domestic consumer, what is known as self-management. For the consumers to do self-management, they require the relevant information pertaining to their consumption patterns. Smart heat meters are already being used to provide this information. However, they are still being under-utilized in terms of their capability. In this research work an Extended Smart Metering Display (ESMD) is proposed; it is based on the interviews conducted with the representatives of smart heat meter manufacturers, District Heating (DH) providers and domestic consumers of DH in the Blekinge county of Sweden. The proposed ESMD was evaluated by the member companies of Swedish District Heating Association and domestic consumers in the workshop conducted for this purpose. The proposed ESMD may help the domestic consumers in monitoring their energy consumption on real-time basis, and improving their energy consumption behavior. It is also suggested that how it can be made more financially viable for the energy consumers and providers during the peak hours, if the proposed system is used.

Keywords consumer behavior measurement, district heating, energy economy, metering display, smart heat meter

1 Introduction

The extended use of technologies in our daily lives makes us dependent on them and it may change/set our behavior. All technological products need some sort of energy to function e.g. vehicles need fuel and home appliances need electricity or gas. On the other hand, the increase in population

leads to more demand of already limited natural energy resources. Moreover, it becomes difficult for government sector to generate and manage these resources economically which may result in higher energy prices [1-4]. Therefore, there is the need of finding either new energy sources or/and to control

the consumption of existing energy resources. One possible solution is to protect and economically use currently available natural resources. It may be achieved by improving existing technology and the related products.

Technological methods are used to improve the generation and distribution of heat and electricity energy. District Heating (DH) is an innovative heating system, becoming popular with the passage of time [5]. Combined Heat and Power (CHP) or DH is a method used in smart grids for the generation and distribution of heat and power energy. It was observed that about two-third of the useful energy was being wasted in the electricity generation process through traditional power plant [6]. DH re-uses this wasted energy for managing the heat energy demands of the power plants, residential, commercial and industrial consumers [7-9]. In the distribution structure of the DH; heat production units, networks of distribution pipes and consumers are its main elements.

It is observed that the residential consumers do not have much self-awareness of economic energy consumption. Sometimes, these consumers have facility through DH available devices to get metering information/services for using DH economically in peak hours. Currently a consumer obtains this information from energy consumption highlights available on meter display or by self-experienced methods. According to [10], the level of energy usage highly depends on the energy consumption behaviors of the consumers living in that environment. However, when the consumers do not get transparent detail of their energy consumption then it be-

comes very difficult for them (despite of their willingness) to improve their behavior for energy conservation.

Currently available residential smart heat meters (Fig. 1) display various technical information for the consumers e.g. the total number of consumed units. The consumers do not know from the display that when and how much energy they consumed in certain duration. This information is not encouraging for the consumers in order to change their energy consumption behaviors. There is also no concept of graphical outputs on these meters for the consumers to inform them about their consumption or to predict future consumption [11]. Consumers cannot analyze that how much energy they have consumed in the previous hour, day, week or month and what was the energy price at that time. In the same way, the display of these meters do not have energy related forecasts for future billing based on current and/or previous energy consumption behaviors of the consumers [11]. Furthermore, the smart heat meters presently provided to the residential consumers of DH do not have any type of energy consumption comparisons on their display for improving their energy consumption behavior.

This research work is an attempt to propose the extended smart metering display for improving the consumers' energy consumption behavior. The new display may achieve it in the following ways:

- providing useful energy consumption feedbacks to the consumers
- facilitating consumers with better energy consumption management in peak hours
- offering consumers and providers with better financial management

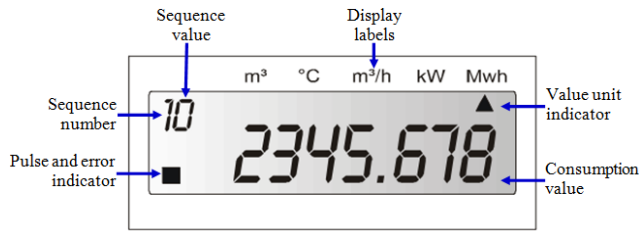


Fig.1. Display of SVM F4 heat meter

2 Research Methodology

Mixed-methods research methodology was adopted in this research; background study and interviews were followed by workshop-based evaluation of the proposed display.

2.1 Interviews

As the research was about measuring daily life utility i.e. DH, which a person can use in different ways round the clock; it was not possible to directly observe the operation of DH providers and the daily activities of DH residential consumers for their energy consumption and saving behaviors in their residences. Therefore, interviews were conducted with the representatives of three different groups i.e. smart heat meter manufactures, DH providers and residential consumers. All the interviewees were selected on the basis of their previous experiences, active participation and the desire for economic usage of DH.

In total, eight interviews were conducted. Out of these eight interviewees, one interviewee was the representative of smart heat meter manufacturers, two interviewees were the representatives of DH providers, and the rest of interviews were the representatives of DH consumers using DH at their homes. The interviewees shared their energy consumption experiences and requirements for extend-

ing the smart metering display. Based on the interviews findings, the guidelines for economic DH consumption and measuring ranges/display patterns related to economic energy consumption behaviors of the consumers were proposed. Financial benefits due to economic energy consumption were also estimated for DH providers. Further, it was also analyzed that how smart metering measurements could be better understood by the consumers.

The representatives of DH providers and the consumers indicated that good consumer behavior may be supported by applying various methods. These methods are lower flow rates of water, billing of peak hours, graphical representation of the energy consumption with comparison of low, medium or high scale, and comparing consumption with similar families in similar area. For some extent, every method is useful in improving energy consumption behaviors of the consumers. However, if to combine all of the above methods in a single tool of smart heat meter with the proposed ESMD, then it can facilitate the consumers to support in improving their energy consumption behaviors effectively by selecting the method of their own choice.

For the purpose of data validity, one of the interviewers repeated the answers in simple words and according to his understanding of what the interviewee said. This practice helped the interviewers to validate their understanding on the spot. Further, member-checking method [12] was used to validate the findings obtained from interviews by sending their copies to the respective interviewees. The interviews findings highlighted the motivations of the interviewees in extending the existing smart metering display for energy economy.

2.2 Workshop-based Evaluation

The quantitative part of this study included the workshop-based evaluation of the proposed ESMD by the member companies of Swedish District Heating Association i.e. Svensk Fjärrvärme. The guidelines indicated in [13] were followed to conduct this workshop. The workshop was conducted in Blekinge Institute of Technology, Karlskrona, Sweden. Participants were the representatives of smart heat meters manufacturers, DH providers and residential consumers. To get the suggestions from consumers, using other types of heating systems in their homes, were also among the participants. Out of total 18 participants, 16 were present physically whereas 2 responded electronically i.e. through email. Firstly, the participants were briefed through multimedia presentation. Later on, the questionnaire was distributed to get the participants' feedback. A set of large sized prints of the proposed ESMD was prepared and displayed on the front wall. The participants were requested to modify the display patterns according to their suitability. This practice started the discussion and questions session to exchange the views and understanding about the proposed ESMD. The participants' feedback indicated the acceptance level and suggestions to further improve the proposed ESMD.

One of the participants (the residential consumer of DH) suggested more frequent or high-resolution feedbacks with the indication of peak energy consumption level of a consumer. In response of it, another participant who was the representative of smart heat meters manufacturers expressed the concern about the high cost of meters for displaying

such feedbacks. According to some other participants, the consumers may be allowed to choose intervals based feedbacks, and different forms of display statistics (numeric and/or graphic) that they want to get on meter display. For the provision of economic energy consumption guidelines, participants indicated the provision of energy saving guidelines with respect to house/apartment type. The provision of such guidelines on the basis of consumers' housing types may motivate them for economic energy consumption but it may also extend the administrative cost for DH providers. Some participants suggested for changing the proposed color combination i.e. red, green and black with red, green and yellow or with red, green and blue combinations. The participants proposed to present green color line for normal consumption, blue color line for consumption slightly above the normal level and red color line for higher consumption levels. The participants also suggested that if the existing color combination is retained then it may be better to change the meaning of the proposed colors; black color for average level, green color for very efficient level and red color for good level of the energy consumptions.

Furthermore, these were also the suggestions by the residential consumer of DH to install heat meters in accessible places in a house like a kitchen etc. and with touch screen facility. On the other hand, the manufacturers of smart heat meters and DH provider representatives regretted to have facility of such installation at the moment. It may take some time to resolve the meter place problem with improvement in the technology.

3 Extended Smart Metering Display (ESMD)

This section highlights the key features of the proposed ESMD after analyzing the feedback obtained from the representatives of three groups during the workshop.

3.1 Physical Interface

Three types of buttons are proposed for smart heat meters. A “Menu” button to move on the main menu, four buttons of arrow keys for navigation, and an “OK” button to finalize the selection.

3.2 Color Scheme

The use of different colors may be helpful to indicate different related events such as peak/off-peak hours, and high, standard and very efficient level of consumption. Therefore, a color scheme is proposed for the proposed ESMD to indicate the events. Table 1 presents the proposed color scheme.

Table 1. Color scheme for the proposed ESMD

Color	Indication
Green	Standard/ normal/ efficient consumption level
	Off-peak hours
Red	Over/ high/ inefficient consumption level
	Peak hours
Black	Very efficient level of consumption

3.3 Instant and Convenient Feedback

Government, energy policy makers or DH providers may set the criteria of the energy consumption levels according to the environmental requirements. The consumers may instantly be able to analyze their energy consumption status at a glance. The proposed display also figures out the total consumption amount with the help of large sized font

to get instant attention of the consumers. Consumers having color-blindness or eye sightedness problems may be served by disability support services like audio alerts and high resolution interfaces available in these future smart heat meters.

Currently, the energy consumption bill presents the total units consumed in Megawatt-hour (MWh) which is a larger unit than Kilowatt-hour (kWh). An ordinary consumer may not know about it or has no interest in the difference between MWh and kWh. The proposed meter displays the energy units in kWh with higher precision. It is proposed to show the amount of bill (e.g. in Swedish Kroner) for quick information.

3.4 Menus/Displays of the Proposed ESMD

Five types of menus/displays are suggested for the proposed ESMD. Table 2 explains the detail of these suggested menus/displays.

It is proposed to have billing display as a default/main display. It shows the current consumption level in a glance (Fig. 2).

Guidance list display provides the guidance of economic energy consumption according to the requirements of peak and off-peak hours' energy consumption behavior. Table 2 lists five associated displays with the guidance list interface such as cleaning guidelines and bath/shower guidelines (Fig. 3).

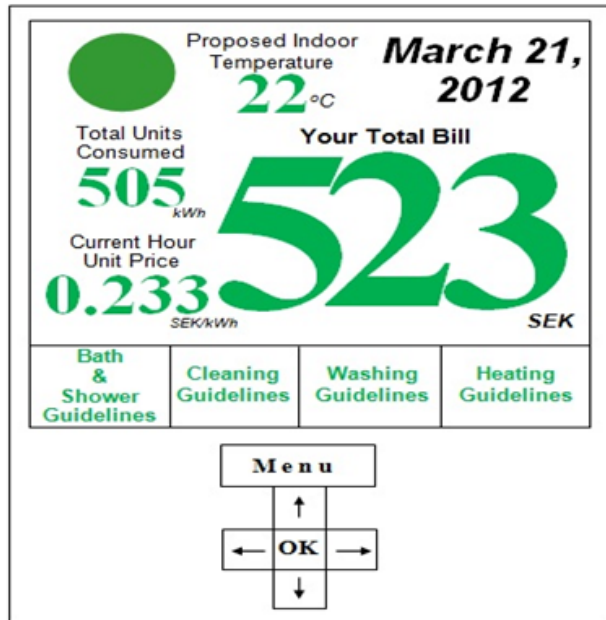


Fig.2. Main/Default display of the proposed ESMD

BATH & SHOWER GUIDELINES
1. Install water-saving nozzle to the shower.
2. Take shower in middle of the day.
3.
4.
5.

Fig.3. Bath and shower display

Feedbacks display of the proposed ESMD provides the energy consumption feedbacks by comparing the current energy consumption level with the previous and predictive future consumption levels (Fig. 4).

PREVIOUS CONSUMPTION	FUTURE CONSUMPTION
Hourly	Hourly
Daily	Daily
Weekly	Weekly
Monthly	Monthly
12-Months' Consumption	Weather Forecast
Compare with Same Month Last Year	
Compare with Similar House	

Fig.4. Feedback menu

The feedback menus display the previous and future consumptions of DH on hourly (Fig. 5), daily (Fig. 6), weekly (Fig. 7) and monthly basis. Besides these displays, the consumers may get their previous 12 months' billing detail, comparison of their last months' consumption with the same month last year and comparison of last months' consumption with families living in same type of houses/apartments. Weather forecast for the next seven days is also part of the proposed ESMD to stimulate the consumers in advance management of their DH consumption behaviors.

Table 2. Color scheme for the proposed ESMD

Sr.	Display Type	Display Name
1	Billing	Default/Main Display
2	Guidance List	Bath and Shower Display
		Cleaning Display
		Washing Display
		Heating Display
		Notifications Display
		Emergency Numbers Display
3	Feedbacks	Previous Consumption Display
		Future Consumption Display
4	Gases Emission	GHGs Emission Display
5	Selection List	Disability Support Display
		Customize Bill Dates Display
		Bill Payment Modes Display

Another interesting display "GHGs Emission Display" (Fig. 8) shows the consumer's participation in energy economy and its impact on reducing GHGs' emissions in percentage form in comparison with the standard level. The value in green color is

the indication of efficient participation level in contrast to red color which indicates inefficient participation level.

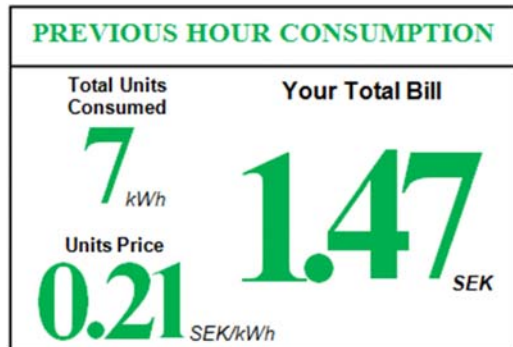


Fig.5. Feedback of previous hour consumption i.e. efficient consumption indication

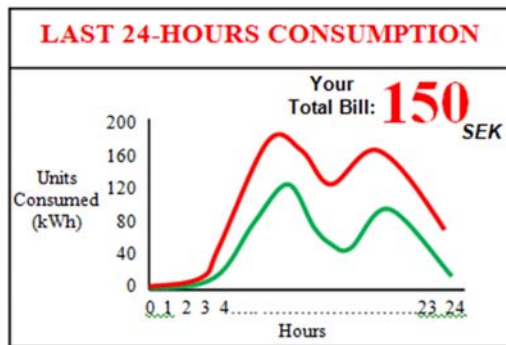


Fig.6. Feedback of previous day consumption (last 24hours) i.e. inefficient consumption indication

NEXT WEEK WEATHER FORECAST						
MON DAY	TUE DAY	WED DAY	THU DAY	FRI DAY	SAT DAY	SUN DAY
8 °C	12 °C	6 °C	8 °C	14 °C	16 °C	18 °C

Fig.7. Weekly weather forecast

3.5 Flexibility in Services

The proposed ESMD provides a display for consumers to select the services according to their special needs. These special needs may be related to disability, bill receiving dates and bill payment

methods.

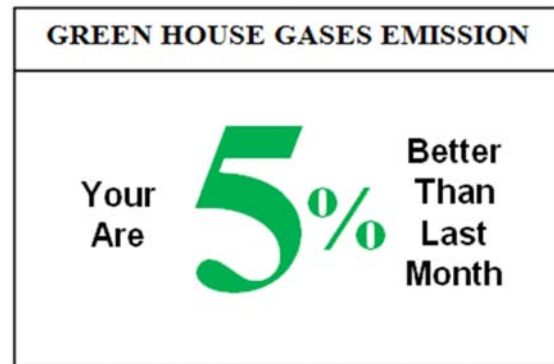


Fig.8. GHGs emission display

3.6 Accessibility

The proposed ESMD provides support for consumers with special needs (Fig. 9). Family members of the consumers having color blindness or eyesight problem may choose the support service (such as audio alerts and high resolution display). Such consumers may also request DH providers for these services.

DISABILITY SUPPORT	
<input checked="" type="radio"/>	Audio alerts.
<input type="radio"/>	High-resolution display.
<input type="radio"/>
<input type="radio"/>
<input type="radio"/>

Fig.9. GHGs emission display

4 Discussion

It is noticed that all the stakeholders are concerned about economic energy consumption. The smart heat meter manufacturers and providers are interested in promoting their business activities. On the other hand, the consumers want to reduce their bills. DH generation plants have fixed capacities of

generating heat energy. On the other hand, heavy investments are required to maintain energy generation capacity of a plant or to extend a plant.

In the bill of DH, the energy providers included two types of charges; the fixed amount of charges for their services and the non-fixed amount of charges for the DH consumption. DH providers get the portion of their fixed charges all the time to keep their system running. On the other side, the billing (non-fixed charges) of consumers is not in the interest of DH providers. Economic DH consumption behavior of the consumers may spare energy in a plant. Providers can sell this spare energy by adding new consumers in the same plant, so in that case the DH providers may get the financial benefit in the form of more fixed charges. Moreover, the reduced energy consumption during peak hours is also financially beneficial for DH providers. In this case, DH providers do not need to use expensive fuels for meeting the peak hours energy demand and it ultimately reduces the DH energy generation cost.

The proposed meter display may help DH providers and manufacturers of the meters to attract the consumers towards the meters display without changing the currently located position of heat meters. Touch screen facility may also help the consumers to operate the meter easily but it may increase the cost of meter and may be difficult for older consumers to operate it. Therefore the touch screen feature is not recommended for the proposed ESMD.

It is found that smart heat meters manufacturers group is concerned with the improvement in smart metering technology for high sale. It will be better to use the same smart heat meter for different

types of houses having appropriate energy consumption guidelines, increased number of notifications and updated billing modes.

5 Conclusion

The ESMD is proposed on the basis of the requirements gathered from the representatives of smart heat meter manufacturers, DH providers and domestic consumers of DH in Blekinge county of Sweden. The proposed display may improve the energy economy by facilitating consumers to manage their high or overlooked energy consumption activities. The improved energy consumption may be achieved through the regular and real time display elements of the proposed ESMD and consumers may be rewarded by having the lower billing. On the other hand, improved energy economy through the proposed ESMD may equally be beneficial for the energy providers. They may economically manage the energy demands particularly in peak hours without extra overhead of fuel-mix or high installation costs on new energy generation plants.

References

- [1] Meehan E. 2010. The smart grid: the smart choice?. In *Proc. Information Security Curriculum Development Conference*, 2010, pp.173-176.
- [2] Soergel S. An economic smart metering pilot implementation using standards-based protocols. In *Proc. IEEE Conference on Innovative Technologies for an Efficient and Reliable Electricity Supply*, 2010, pp. 216-219.
- [3] Choi T S, Kyung R K, Seong C P, Young S J,

- Yong T Y, Sang K I. Analysis of energy savings using smart metering system and in-home display). In *Proc. Transmission Distribution Conference Exposition: Asia and Pacific*, 2009, pp. 1-4.
- [4] Weiss M, Guinard D. Increasing energy awareness through web-enabled power outlets.” In *Proc 9th International Conference on Mobile and Ubiquitous Multimedia*, 2010, pp. 1-10.
- [5] Mahapatra K, Gustavsson L. An adopter-centric approach to analyze the diffusion patterns of innovative residential heating systems in Sweden. *Energy Policy*, 2008, 36(2): 577-590.
- [6] Tsung, Johan. METRIMA-SVM F4 calculator (manual). Metrima AB, 2000.
- [7] Wernstedt F, Davidsson P, Johansson c. Demand side management in district heating systems. In *Proc 6th International Joint Conference on Autonomous Agents and Multiagent Systems*, 2007, pp. 1383-1389.
- [8] Wu D W, Wang R Z. Combined cooling, heating and power: a review. *Progress in Energy and Combustion Science*, 2006, 32 (5-6): 459-495.
- [9] Johansson C. Towards intelligent district heating. PhD Thesis, 2010, Blekinge Institute of Technology, Sweden.
- [10] Parker D, Hoak D, Meier A, Brown R. How much energy are we using? Potential of residential energy demand feedback devices. In *Proc Summer Study on Energy Efficiency in Buildings*, 2006.
- [11] Environmental and Energy Study Institute. District energy: essential infrastructure for energy-efficient communities. Accessed July 19, 2015.
- [12] Steinar Kvale. Interviews: An Introduction to Qualitative Research Interviewing. Sage Publications, 1996.
- [13] Al Rollins. How to conduct a workshop to integrate gender considerations into development planning. ERIC, 1992.

Classifying facial expressions using DWT, Moments and Genetic algorithm

M. Mahadevi ¹, Dr. C. P. Sumathi ²

¹ Research Scholar (M.S.University) & Asst. Professor
SDNB Vaishnav College for women, Chennai, Tamilnadu, India

² Associate Professor & Head, Department of computer science,
SDNB Vaishnav College for women, Chennai, Tamilnadu, India.

Abstract:

Facial expressions are the actions of the thoughts that arise in a mind. Such expressions are categorized as simple basic and complex expressions which are a mixture of two or more expressions. This research focuses on identifying the basic expressions and classifying them based on Naïve Bayes classifier. The database considered for the research is Japanese Female Facial Expression (JAFPE) consisting seven expressions happy, sad, disgust, fear, angry, neutral and surprise. The image is pre-processed using Discrete Wavelet Transform (DWT) and created a feature set containing spatial statistical features of the facial parts and moments of the DWT image. The features were selected using genetic algorithm and classified the database using Naïve Bayes classification to acquire an overall accuracy rate of 92.5%

Keywords: Spatial Statistical features, DWT, Genetic algorithm, Naïve Bayes

I. INTRODUCTION

Facial expressions are the main form of non-verbal communication during any human interaction. Facial expressions recognition has its applications in the field of medicine, education, online communication, personal interviews and crime interrogation reports. Such being the importance of the expression recognition, the research on classifying the expressions is based on different techniques and phases. Kezheng Lin et.al has applied three-dimensional space of false geodesic distance to classify the expressions [1]. Vikas Maheshkar et.al has applied discrete cosine transform and using the energy distribution has classified the expressions [2]. Researchers have applied several techniques like Local Directional Pattern [3], Log-Gabor Filters with SVM [4], and Geometrical deformation feature vectors for SVM [5]. G.U.Kharat has applied three different feature extraction techniques and classified the expressions using SVM [6], Poonam Dhankar et.al has applied Gabor filters and Eigenvectors [7], Ramchand Hablani et.al has applied Local Binary Patterns for important facial parts[8], Shilpa Choudhary et.al has applied hybrid feature extraction and Adaboost classifier[9] M.Mahadevi et.al has applied template matching on mouth detection using genetic algorithm to classify the expressions[10] , and they have used JAFPE[11] database for the research. Still classifying the facial expressions has been the most challenging problem in the field of research. So, the proposed work has applied Statistical feature extraction to the facial parts and Moment extraction for the DWT image to create a feature vector of 17 features. A genetic algorithm is applied to select the best features, and Naive Bayes classifier is used for classifying the expressions. This work has also compared the existing methods applied on the database JAFPE with the proposed work.

II. PROPOSED METHOD

The proposed work of facial expression recognition system has different levels. The Block diagram of the proposed system is given in Fig 1.

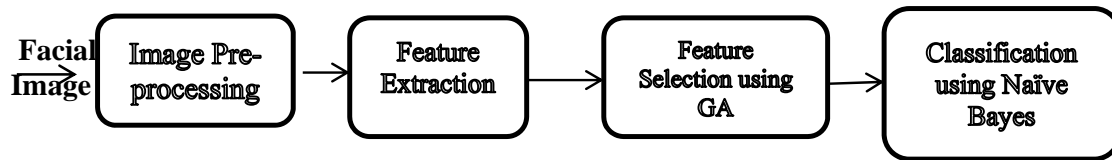


Fig 1: Block diagram of the proposed system.

III. IMAGE PRE-PROCESSING:

Image pre-processing is a processing step wherein the captured images are transformed to another image which may be a segmented image or transformed frequency domain image or an enhanced image required by the kind of research and techniques. Two image pre-processing techniques that are applied in this research work are image enhancement using a genetic algorithm and creating a transformed image using Discrete Wavelet transform. The input image is filtered using an optimal filter [10] that was created using a genetic algorithm. The resulted image is converted to a binary image using a suitable threshold value and is manually cropped to obtain the face boundary and is resized to $128 * 128$ for further processing. This resulted image referred as input image 1 is given in fig 2. The same input is passed through Discrete Wavelet Transform to create a transformed image referred as input image 2.



Fig.2: Input image 1- Enhanced image

A. Discrete Wavelet Transform:

Image transforms allow converting the image from the time domain to frequency domain to enhance the feature extraction stage of image classification. Wavelet transform can decompose a signal into sub-bands with low frequency (approximate components) which are consistent with characteristics of a signal and sub-bands with high frequency (detail components) which are related to noise and disturbance. The two level discrete wavelet transform resulted in four components the low-frequency component, Horizontal components details, vertical component details and the diagonal components details. If the size of the input image is $n * m$ then after the first filtering the size of the image gets reduced to $n * m/2$ and by applying further filtering, the Low component image further gets reduced to $n/2 * m/2$ which is the two level Discrete wavelet transforms. The

sub-band decomposition of the input image LL, LH, HL and HH is given in fig 3, and the corresponding input image decomposition is given in fig 4.

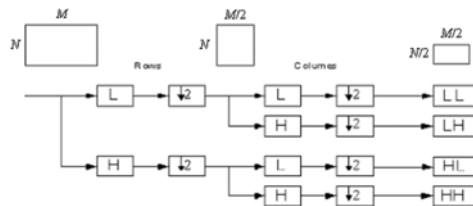


Fig.3: Decomposition of Input image of size $n \times m$

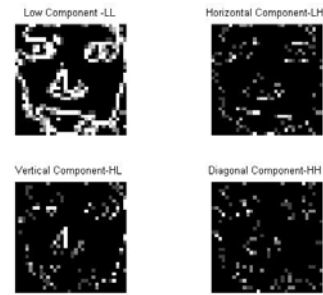


Fig.4: DWT transformed image

LL is an approximate image of the input image, and it is a low-frequency subband. LH subband extract the horizontal features of the original image, HL subband gives vertical features, and HH subband gives diagonal features. The LL component of the decomposition carries the most useful information and moments on the LL component are the features that are extracted. The first image in Fig.4 is LL component image referred input image 2.

IV. FEATURE EXTRACTION

Feature extraction is an important phase which generates features to help in the task of object classification. Two kinds of features are extracted from this work. A set of spatial statistical features is extracted from the facial parts of the input image 1 and moment features for the input image 2- DWT transformed image. A combined feature set is created and is passed through the feature selection stage.

A. Spatial Statistical features:

To create the statistical features set, the input image 1 is divided into upper face region and lower face region using horizontal and vertical projections. Horizontal projection is the sum of the row pixels, and vertical projection is the sum of the column pixels. The sum total value of the pixels of the cropped face image helps to identify the boundary of the face. Using the peak value of the horizontal projection the cropped face image is divided into lower and upper face. The upper face image is further divided into left and right face regions based on the peak value of the vertical projection of the upper face.

An edge contains useful information about the object. So edges are used to measure the size of the objects in an image. Sobel edge operator is applied to left upper face region, right upper face region and lower face region as in fig 5, fig 6 and fig 7.



Fig: 5
Left upper face



Fig: 6
Right upper face



Fig: 7
Lower face

The cropped face is divided into three facial parts as left upper, right upper and lower face regions. Spatial statistical features mean, standard deviation, entropy were used to create the feature data set for the input image1. So a feature vector with 9 features (3 from each facial part) is extracted.

Mean: Mean of a region calculates the average value of the pixels among the range of pixel values in that region and is given by the equation.

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

where X_1, X_2, \dots, X_i represent pixel values and n is a total number of pixels of that region.

Standard deviation: Standard deviation of a region is a statistical measure that quantifies the amount of dispersion of the pixel values around the mean. It is given by the equation.

$$\sigma = \sqrt{(1/n) \sum_{i=1}^n (x_i - \mu)^2} \quad (2)$$

where μ is the mean of the facial part region, and N is a total number of pixels of that region.

Entropy: Entropy of a region is a statistical measure of randomness that can be used to characterize the texture of the input image and is given by the equation.

$$\text{Entropy} = - \sum_j P_j \log_2 P_j \quad (3)$$

where p_j is the probability, that the difference between 2 adjacent pixels is equal to i .

The feature vector of the input image 1 and let it be named as feature vector 1 as in table 1.

TABLE .1 FEATURE VECTOR 1

Feature Vector 1
Leftmean
Leftstd
leftent
Rightmean
Rightstd
Rightent
Lowmean
Lowstd
lowent

Where leftmean, leftstd, leftent and rightmean, rightstd and rightent are the mean, standard deviation and entropy of the left facial region and right facial region. Lowmean, lowstd and lowent are mean, standard deviation and entropy of the lower face region.

B. Moment Features

The weighted average of the image pixels' is the image moment. Based on normalized central moments, Hu [12] introduced seven-moment invariants. The Hu Moments helps to identify the shape of an object in an image. Usually, the moments are calculated using the outline of the object so in this work also the moments are calculated for the input image 2. The input image is low component image wherein the expressions changes are more visible on the edges of the image which help to identify the shape of the objects and their

intensity levels. The appearance of the face differs from expressions, and it changes the shape of the objects present in the face. The moments of the images exhibiting expressions are calculated to form an 8-element feature vector 2 as in table 2.

TABLE.2 FEATURE VECTOR 2

M1	M2	M3	M4	M5	M7	M8
----	----	----	----	----	----	----

where M1, M2, M3, M4, M5, M6, M7 and M8 are the moments in terms of central moments. The moment m8 was proposed by J. Flusser and T. Suk [13], and they have proved that m8 is a third order independent moment invariant to be more useful than the m3 which is shape descriptor moments.

C. Combined Feature vector

A combined feature vector is created using the spatial statistical features (Feature vector1) and moments (Feature vector 2) to form a 17-element feature vector as in table 3 for the JAFFE database, and the dataset is further subjected to feature selection process to select the best features which will improve the performance of the classifier. The combined feature vector is tabulated with numbers from 1 to 17 in table.3

TABLE.3 COMBINED FEATURE VECTOR

Leftmean	1
Leftstd	2
Leftstd	3
Rightmean	4
Rightstd	5
Rightent	6
Lowmean	7
Lowstd	8
Lowent	9
M1	10
M2	11
M3	12
M4	13
M5	14
M6	15
M7	16
M8	17

The sample dataset values are given in fig 8.

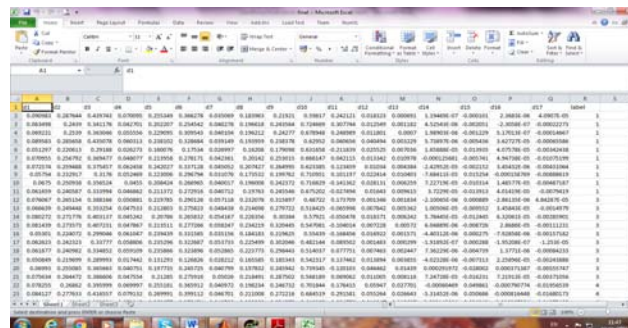


Fig.8: Sample dataset

V. FEATURE SELECTION & CLASSIFICATION

The algorithm uses Naïve Bayes classifier for classification and genetic algorithm for feature selection. As an initial step of classification, a Naïve Bayes classifier object was created using the training data set, based on which classification accuracy is calculated. To improve the accuracy of the algorithm feature selection was performed using a genetic algorithm.

A. Naïve Bayes classifier

Naïve Bayes classifiers [14] are probabilistic classifiers based on Bayes theorem with independence between the features. It is stated that each of the attributes contribute equally to the classification problem. By analyzing the contribution of each of the independent attribute a conditional probability is determined and a classification is made by combining the impact that the different probabilities have on the prediction to be made.

For a given training data the Naïve Bayes algorithm finds the prior probability for each class by counting how often each that class occurs in the training data. For each attribute x_i , the number of occurrences of each attribute value is counted to determine $P(x_i)$. Then the probability $P(x_i/c_j)$ can be estimated based on the occurrence of value occurs in the class in the training data. While classifying a target tuple, the conditional and prior probabilities generated from the training set are used to make the prediction. Estimation of a tuple 't' which has p independent attribute values $\{x_{i1}, x_{i2}, x_{i3}, \dots\}$ is given by the equation 4.

$$P\left(\frac{t_i}{c_j}\right) = \prod_{k=1}^p P\left(\frac{x_{ik}}{c_j}\right) \quad (4)$$

In this algorithm, a Naïve Bayes classifier object is created using a training dataset with 140 rows and 17 column values wherein the row contributes the observations from each image, and the columns contribute the features of the images. With the predicted class labels from the classifier object, the algorithm then classifies the test data based on the largest posterior probability.

B. Feature selection using genetic algorithm

Feature selection is required when there is a number of features or when there are so many independent attributes. In this work since each of the attribute values are independent and contributes equally to the classification problem, feature selection is needed for best classification rate with minimum features. Hence, a genetic algorithm is used for feature selection.

Genetic algorithm [15] is an evolutionary algorithm that generates solutions to optimization problems using techniques such as mutation, crossover, and selection. During each iteration several generations are generated with a set of candidate solutions. Those solutions are called as the population. The evolution starts with randomly generated individuals and is an iterative process. The fitness of every individual in the population is evaluated for each generation, and that is the value of the objective function in the optimization problem being solved. When the population satisfies the fitness, it passed through the next phase and based on mutation or crossover they are selected for the next iteration. The best fitness from the generations converges to an optimal solution for the problem. The convergence is realized either when maximum generations are generated, or when the satisfied fitness value is reached.

The fitness function defined for this research is to minimize the Root Mean Square Deviation (RMSD) of the Classification with the randomly selected features. The RMSD represents the sample standard deviation of the differences between predicted values and observed values [16]. The fitness function is given in equation 5

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^n (\hat{y} - y)^2}{n}} \quad (5)$$

Where \hat{y} the predicted value and y is the original value and 'n' is the number of values.

As RMSD gets converges to a minimum value, the features that give the best accuracy are selected. The population is a binary string length of 17 features. Seed chromosome is initialized randomly and gets into generation, and a typical chromosome structure is given in table 4.

TABLE 4: STRUCTURE OF A CHROMOSOME

1	1	0	0	0	0	1	1	0	0	1	0	1	0	1	0	0
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

The ones in the chromosome indicate that the corresponding feature will be selected and will be used for classification and RMSD is calculated for each stall generation. When the specified limit stall generation is reached, the best fitness is sent to the next generation run. Repeating the generation's up to a maximum value, the best fitness is plotted, and the respective features are selected. After the genetic algorithm reaches its stopping criteria, the selected features of the test data are classified using Naïve Bayes classifier.

VI. RESULTS AND DISCUSSION

The proposed algorithm for facial expression classification used JAFFE database which consists of 213 images expressing 6 basic expressions like happy, sad, angry, surprise, fear, disgust and one neutral expression. The algorithm classifies the basic expressions by excluding the neutral expressions. Training is done in two phases, and the dataset is divided into two sets, one consisting of 140 images and a second set consisting of 80 images. In the first phase, 140 images are trained, and features were selected and tested on the second set of images. In the second phase using 'resubstitution' method, the second set of images were trained and tested. The work classifies the six expressions without considering the neutral expression.

The Genetic algorithm parameters for the training are given in table 5.

TABLE 5: GENETIC ALGORITHM PARAMETERS

Parameters	Values
Population Type	Bit String
Population Size	10
Generations	50
Stall Generation Limit	10
Mutation	Uniform
Selection	Roulette Wheel
Cross Over	Cross over Arithmetic

The convergence of each of the training data set is plotted in the fig 9 and fig 10.

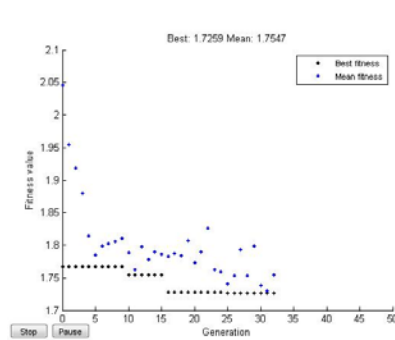


Fig 9: Training Data Size: 140
RMSD Error rate: 1.7259

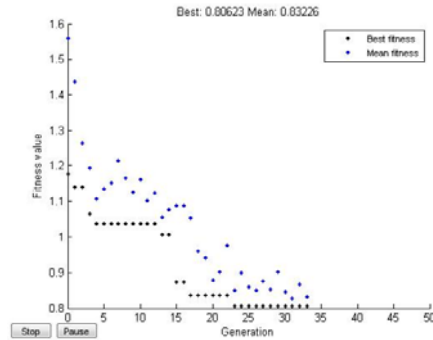


Fig 10: Training Data Size: 80
RMSD Error rate: 0.80623

The selected features for each dataset are given in table 6.

TABLE 6: FEATURES SELECTED USING GENETIC ALGORITHM

Training Dataset	Features selected	RMSD
140 images	1,2,6,7,9,10,12,13,14,16,17	1.7259
80 images	1,3,5,8,10,12,13,14,15,16	0.8062

From the table 6, it is very clear that the RMSD is less when training the algorithm with data set of 80 images. Both the selected features set are used to classify the expressions using Naïve Bayes classifier. The corresponding confusion matrix for 140 images is shown in table 7. The confusion matrix for 80 images is shown in table 8. Performance accuracy for all the test dataset sizes is tabulated in table 9.

TABLE: 7 CONFUSION MATRIX FOR 80 IMAGES (TRAINING: 140 AND TEST SET: 80)

	Sad	Happy	Disgust	Surprise	Fear	Angry
Sad	8				2	
Happy	2	12			1	1
Disgust	1	3	9			1
Surprise		1		13		
Fear					11	
Angry	1					14

TABLE: 8 CONFUSION MATRIX FOR 80 IMAGES (TRAINING AND TEST SET: 80)

	Sad	Happy	Disgust	Surprise	Fear	Angry
Sad	10					
Happy	1	15				
Disgust	1	3	10			
Surprise				14		
Fear					11	
Angry	1					14

TABLE: 9 PERFORMANCE ACCURACY FOR DIFFERENT DATASET SIZES

Training Dataset	Test Dataset	Features selected	Accuracy
140 images	80 images	1,2,6,7,9,10,12,13,14,16,17	83.75%
80 images	80 images	1,3,5,8,10,12,13,14,15,16	92.5%
80 images	80 images	All 17 Features	88.75%

From the table 3, 4 and 5 it is clear that the dataset with 80 images uses a minimum number of features and has better recognition rate than dataset of 140 images. Table 6 indicates that the ‘restitution’ method of selecting the test data shows a remarkable accuracy rate than the ‘hold out’ method. The overall accuracy of the algorithm for 80 images is 92.5%.

A. Comparison with other methods

Poonam Dhankar and Neha Sahu have used Gabor filter and Eigen Vector [7] on JAFFE database with different dataset sizes of 10,20,30,40 and 50 images and reached a mean accuracy of 93.15%. Ramchand Hablani et.al has used Local Binary Patterns on facial parts [8] with an accuracy of 73.61 % for person independent classification and 94.44 % for person dependent classification. M.Mahadevi and C.P. Sumathi have used mouth detection and template matching using genetic algorithm [10] on 50 images to reach an accuracy of 94%.

The proposed algorithm is a person independent classification on 80 images with the best features selection and has a better accuracy rate of 92.5%. The accuracy analysis of different methods of the proposed methodology is tabulated in Table 10.

TABLE: 10 COMPARISON WITH OTHER METHODS

Methods	Dataset size & Constraint	Accuracy
Gabor Filters + Eigen Vector[7]	50 images	93.15%
Local Binary Patterns+Template matching[8]	213 images(Person Independent)	73.61%
Mouth Detection +Template matching[10]	50 images	94%
Proposed methodology DWT + Statistical & Moment features + Naïve bayes	80 images	92.5%

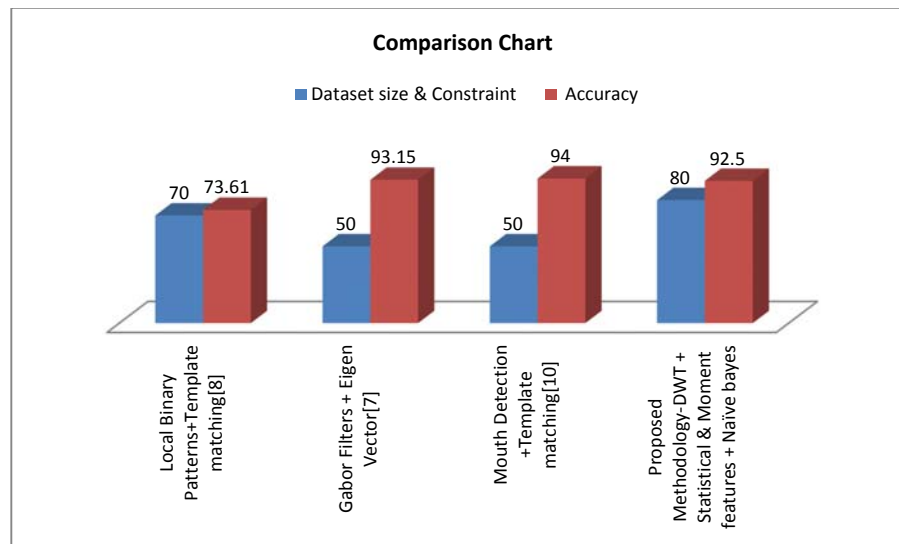


Fig 11. Comparison of different methods with proposed method

Fig. 11 clearly shows that the proposed method has a better accuracy in terms of the dataset size and constraint on the classification than the other methods.

VII. CONCLUSION

This paper has developed a combined feature dataset with spatial statistical features mean, standard deviation and entropy of the facial parts left eye, right eye, and lower face, DWT based moment features of the low-frequency component. Features are optimally selected using a genetic algorithm with RMSD as the fitness function. The feature selection was given two sets of training data comprising of 140 images and 80 images to create two different features set. Based on the RMSD the features are selected and classified using Naïve Bayes classifier. Using the Hold out method around 57% of training data (140 images) are considered for testing to result in an accuracy of 83.75%. Using Resubstitution method, 80 images are tested to yield an accuracy of 92.5%. So, the algorithm resulted in an overall accuracy of 92.5%. The proposed methodology is compared with other methods to show that the proposed has a better accuracy than the existing methods.

VIII. REFERENCES

- [1] Kezheng Lin, Weiyue Cheng and Jingtian Li, "Facial Expression Recognition Based on Geometric Features and Geodesic Distance", International Journal of Signal Processing, Image Processing and Pattern Recognition Vol.7, No.1 (2014), pp.323-330 <http://dx.doi.org/10.14257/ijsp.2014.7.1.30>
- [2] Vikas Maheshkar, Sushila Kamble, Suneeta Agarwal and Vinay Kumar Srivastava, "Feature Image Generation Using Low, Mid and High-Frequency Regions For Face Recognition", International Journal of Multimedia & Its Applications (IJMA) Vol.4, No.1, February 2012.
- [3] Adin Ramirez Rivera, Jorge Rojas Castillo, Oksam Cha "Local Directional Number Pattern for face analysis", IEEE Transactions on Image Processing, Vol.22, No.5, 2013.
- [4] Jeny George, "Facial Expression Recognition from color images using log-Gabor Filters", International Journal of Current Engineering and Technology, Vol.5, No.2 April 2015.
- [5] Rajesh Patil, Vineet Sahula, A.S.Mandal, "Features Classification using Geometrical Deformation Feature Vector of SVM and Active Appearance algorithm for Automatic facial expression recognition", Machine Vision and Applications, April 2014 Impact Factor: 1.35 · DOI: 10.1007/s00138-014-0594-5.
- [6] G.U.Kharat, Dr. S.V.Dudul "Human Emotion Recognition System Using Optimally Designed SVM With Different Facial Feature Extraction Techniques", WSEAS Transactions On Computers, Vol. 7, No.6 June 2008.
- [7] Poonam Dhankar, Neha Sahu "A Novel algorithm for Human Facial expression recognition using Gabor Filters and Eigen Vector", International Journal Of Computer Applications, Vol.97, No.7, July 2014.
- [8] Ramchand Hablani, Narendra Chaudhary, Sanjay Tanwani, "Recognition of facial expressions using Local Binary pattern of important facial parts", International Journal of Image Processing, Vol.7, No. 2, 2013.
- [9] Shilpa Choudhary, Kamlesh Lakhwani, Shubhlakshmi Agarwal, "An efficient hybrid technique of feature extraction for facial expression recognition using AdaBoost Classifier", International Journal of Engineering research and Technology, Vol.1, No.8, October 2012.ISSN:2278-0181.
- [10] M.Mahadevi, Dr.C.P.Sumathi, "Detection of Mouth for Facial Expression classification using Genetic Algorithm", International Journal of Applied Engineering Research ISSN 0973-4562 Vol.10, No 16 (2015).
- [11] <http://www.kasrl.org/jaffe.html>
- [12] M. K. Hu, "Visual Pattern Recognition by Moment Invariants", IRE Trans. Info. Theory, vol. IT-8, pp.179-187, 1962
- [13] J. Flusser and T. Suk, "Rotation Moment Invariants for Recognition of Symmetric Objects", IEEE Trans. Image Proc., vol. 15, pp. 3784-3790, 2006
- [14] https://en.wikipedia.org/wiki/Naive_Bayes_classifier.
- [15] https://en.wikipedia.org/wiki/Genetic_algorithm
- [16] https://en.wikipedia.org/wiki/Root-mean-square_deviation

A Modern Approach to Integrate Database Queries for Searching E-Commerce Product

Ahmad Tasnim Siddiqui (Corresponding author)
College of Computers and Information Technology, Taif University
Taif, Saudi Arabia

Mohd. Muntjir
College of Computers and Information Technology, Taif University
Taif, Saudi Arabia

Abstract- E-commerce refers to the utilization of electronic data transmission for enhancing business processes and implementing business strategies. Explicit components of e-commerce include providing after-sales services, promoting services/products to services, processing payment, engaging in transaction processes, identifying customer's needs, processing payment and creating services/products. In recent times, the use of e-commerce has become too common among the people. However, the growing demand of e-commerce sites have made essential for the databases to support direct querying of the Web page. This re-search aims to explore and evaluate the integration of database queries and their uses in searching of electronic commerce products. It has been analyzed that e-commerce is one of the most outstanding trends, which have been emerged in the commerce world, for the last decades. Therefore, this study was undertaken to ex-amine the benefits of integrating database queries with e-commerce product searches. The findings of this study suggested that database queries are extremely valuable for e-commerce sites as they make product searches simpler and accurate. In this context, the approach of integrating database queries is found to be the most suitable and satisfactory, as it simplifies the searching of e-commerce products.

Keywords: E-commerce product search, e-commerce, query optimization, business processes, Query integration.

I. INTRODUCTION

The ability of the e-commerce to enable the users to search conveniently for the products in databases is critical to its success. Even though data base queries are considered as the most effective method to access the product database of the e-commerce sites, no significant amount of researches have enlightened the benefits of integrating database queries with e-commerce product searches (Agrawal et al., 2001). In this paper, we have highlighted how e-commerce searches over structured product entities can be optimized by keyword queries such as "iphone 6". However, one major challenge of using database queries for e-commerce product searches is the language gap between the specifications of the products in the databases and the keyword utilized by the people in the search queries (Vander Meer et al., 2012). Google style search box is the most extensively used web interface where the submitted queries neither attribute unit or names.

The intention of this paper is to draw attention towards database queries as well as their use in e-commerce products searches. According to Li and Karahanna (2012), electronic commerce can be explained as the trade of services and/or products, by using internet. E-commerce database queries can be understand as one of the most important database operations, which are totally based on the relational model. The relational models were established by Codd. The term relation is used here in its accepted mathematical sense. Given sets S_1, S_2, \dots, S_n , (not necessarily distinct), R is a relation on these n sets if it is a set of n tuples each of which has its first element

from S1, its second element from S2 , and so on. It is important to notice that evolution of query form is one of the most effective and integrated user interfaces, which are being used for querying the databases for various applications (Vander Meer, et.al, 2012). An example of the query based searching can be the user searching for a laptop on e-commerce sites with a processor rated around 2.4GHz and around 4 GB of internal memory, can make a query by inserting laptop 2.4 4gb (Agrawal et al., 2001). In this case, it has been assumed that if the e-commerce databases get a close match on numbers then it is possible to get an accurate match for the attribute names. The integration of database queries provides a unified access to various e-commerce search engines (Chiu et al., 2014). These optimized database queries are of great importance since they allow the users to search and compare the products and services of different brands from various sites. To avoid the bottleneck situation of database server and web server we should use the optimized and scalable queries.

II. AIMS AND OBJECTIVES

This research aims to explore integration and utilize of database queries in e-commerce product searches. The term e-commerce refers to the selling, buying and commerce performed electronically i.e. by the web based services. With the advancements in internet and emergence of new technologies, e-commerce is turning out to be highly popular among people due to its benefits (Poggi, et al., 2012). The concept of e-commerce is the center of majority of the discussion; however, the concept lacks an inclusive definition which can be accepted widely (Chiu et al., 2014).

Generally, e-commerce has been defined as “the tools which enhances the relationship of an organization with its stakeholders”. The feature common in both the definitions is the significance of customer relationships in terms of their maintenance and establishment (Grandon et al., 2011). Likewise, the development of an e-commerce site can have a significant impact on the transaction costs (Xiao & Benbasat, 2011). For example, through these websites the organizations can make transactions at a relatively lower time and effort. However, there are certain short comings of transaction based websites (Grandon et al., 2011).

III. SIGNIFICANCE OF RESEARCH

According to Das-Sarma, et.al, (2014), due to the technological improvements have taken place, there is significant changes in the lives of people, in terms of daily routine works. In this regard, electronic commerce can be measured as one of the most prominent developments, which have been occurred in retailing industry. In the studies of Endrullis, et.al, (2012) it has been documented; e-commerce has been established very fast, across the world. According to the Li and Karahanna (2012), database queries are found to be one of the integrated and important methods, which assist users in searching required products or services, in extensively huge and sophisticated databases. The technique of database queries provides in instant and quick product output. In this manner, it can be stated that the integration of database queries in e-commerce product searches is one of the ultimate advantages towards integrated and viable trading activities.

IV. REVIEW OF LITERATURE

A. E-commerce Concept and Characteristics

According to Li and Karahanna (2012), e-commerce can be explained as number of activities, including exchanging, selling, and buying of products, services, and information, using computer system networks, primarily

internet. It is important to observe that the term “commerce”, is usually referred to the transactions, which are conducted among business entities.



Figure 1. E-commerce (Source: <http://www.episerver.com/e-commerce/>)

Various practitioners as well as academics have proposed multiple definitions of e-commerce. In simple terms, e-commerce corresponds to online shopping. In terms of business, the definition of e-commerce is not limited to only selling or buying products through internet but also encompasses various processes such collaborating with clients and customers, providing customer service before and after a sale (Lu et al., 2010). The definition of e-commerce proposed by Grandon et al., (2011), summed these processes as a broad array of activities and up and down the value added chain.

Schneider & Perry (2001) defined e-commerce as “the utilization of electronic data transmission in order to enhance business processes and implement business strategies”. The term business processes in this definition corresponds to the activities in which the business engage as they attain explicit aspects of commerce. As given by Grandon et al., (2011) explicit components of commerce in relation to supplier include providing after sales services, promoting services/products to services, processing payment, engaging in transaction processes, identifying customer’s needs, processing payment and creating services/products. Lu et al., (2010) argued that all these activities or constituents of commerce can be successfully achieved by the means of electronic commerce technologies. However, some of the processes related to business utilize traditional commerce activities in a more effective manner (Lu et al., 2010).

B. Classification of E-Commerce

For the purpose of this study, the classification of e-commerce is based upon the business format and business focus. In terms of business focus, the type of business focus is identified by the means of type of buyer, which can be either business clients or end product consumers (Xiao & Benbasat, 2011). In situations, when the buyer is the end consumer, the e-commerce is termed as business-to-consumer e-commerce (B2C). Some websites which are viewed as business-to-consumer e-commerce are eBay.com, Barnesandnoble.com and Amazon.com. On the other hand, when the purchaser is business client or organization, the e-commerce trade is termed as business-to-business e-commerce (B2B) (Fang, 2011).

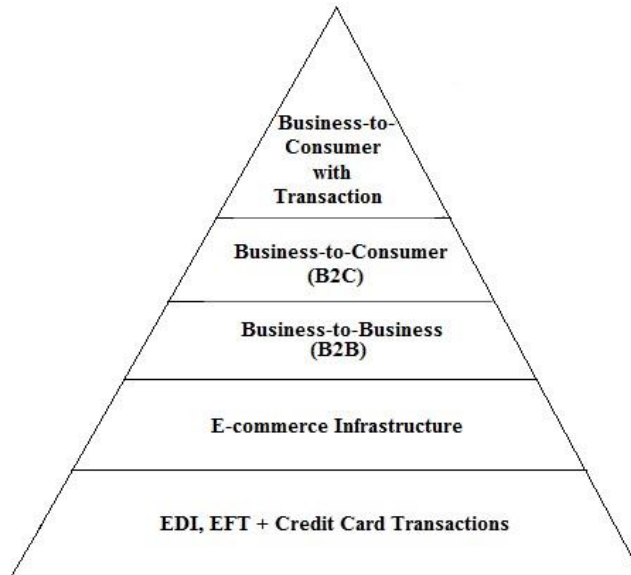


Figure 2. Classification of E-commerce (Source: Fang, 2011)

This model can undertake a range of forms, for instance, there are web based platforms, B2B e-market places and B2B storefronts which gather different retailers and sellers in virtual environment. Some of the chief difference among B2C and B2B are listed below:

- In business-to-business e-commerce, the buyers can purchase larger quantities of desired products.
- In business-to-business e-commerce (B2B), the mode of payment is characterized by purchase order whereas in business to consumer e-commerce (B2C), all the payments are made by credit card regardless of the purchase order.
- In business-to-business e-commerce negotiations are more common and reporting is done by more advanced method (Hinz et al., 2011).
- In business-to-business ecommerce, relationships are considered as extremely crucial.
- In business-to-business e-commerce switching cost is relatively higher.

The figure-2 demonstrates the classification of e-commerce. The virtual teams which are functioning under business-to-business e-commerce are Dell.com and Paper Exchange.com. Although Dell.com also sells its services and products to the consumers, its chief transaction value is achieved through business clients (Hinz et al., 2011). However, it is termed as business-to-consumer e-commerce if it receives its maximum sales from end consumers. In contrast, if the main sale revenue is generated from the business clients, Dell.com is regarded as business-to-business e-commerce.

The e-commerce can also be classified according to its business format. In situations when the chief revenues of e-commerce are generated to online medium, it is termed as an online-dominated channel e-commerce (Casterella & Vijayarathy, 2013). On the other hand, if the revenues are gained from a non-internet medium, the e-commerce is regarded as a traditional dominated channel e-commerce. Various online sellers such as Amazon.com, eBay.com and Dell.com by selling their products and services online have promoted the concept of e-commerce (Hinz et al., 2011). Subsequently, few auction stores had decided to sell their products through auctions rather than fixed priced deals.

The increasing popularity of online auctions had fostered the creation of various online auction forums such as eBay.com (Geerts & Poggi, 2010). In the recent times, the online auctions have gained so much of popularity that all most all the online stores have initiated a section of auction in which, they sell their products at bargained rates along with fixed price products (Xiao & Benbasat, 2011). All and all, the concept of e-commerce has turned out to be so popular that every service or product desired by the individuals can be found on with negotiable as well as fixed cost.

V. DATABASE QUERIES AND INTEGRATION

Database queries enable the users to retrieve pertinent data from a data. This implies that without searching the entire table, the users can point out the categories of data which can be sought further (Geerts & Poggi, 2010). In addition, database queries also enable the users to merge multiple tables, for instance, if a user is dealing with two tables named as invoices and customers, they can utilize database query to merge the contents of two tables. Subsequently, if a user runs this query, he can attain results which illustrate the name of the customers according to their invoices. It is eminent to mention here that a database query only point out data and does not deals with its storage (Vander Meer et al., 2012). Some of the potential benefits of database query are listed below:

- Merge data from various data sources
- Allows the user to select fields from diversified resources and specify them accordingly
- Identify the records that match the criteria set by the users

a. Query Reasoning and Expanding Module

In contrast to traditional query systems, the key attributes of semantic query include expansion of reasoning functions into the queries of user during the querying process (Agrawal et al., 2001). In situations when the semantic information is recommended, the visited records are initially inquired to swiftly identify the interested goods that are allied with the desires of the consumers.

b. Query Breaking Module

It is essential to breakdown selected queries further into atomic queries in order to minimize the complexity allied with searching (Vander Meer et al., 2012). This process is mainly carried out by query breaking module. One issue that arises in this module and needs to be resolved is to determine the most appropriate LECO and position at their defined locations.

c. Integration of Database Queries with E-Commerce

It is widely recognized that the technology of e-commerce is spreading its horizons rapidly as the buyers are increasingly switching their choices to online stores and markets for purchasing personal care items, clothing and other products or services (Poggi et al., 2012). It is eminent to note here that 6 % of all the revenues are generated through online stores. The midsize sellers are particularly quite interested in e-commerce trading since it allows them to expand their market share and compete with large sized firms. Integrating database queries with e-commerce product searches has numerous benefits (Chiu et al., 2014). For instance, database queries eliminate ambiguity while performing e-commerce product searches. Apart from benefiting e-commerce trading, database queries can be integrated with numerous applications such as custom built apps, personalized marketing apps, e-commerce application, CRM systems and ERP systems.

Database queries when integrated to e-commerce product searches ascertain durability, isolation, consistency and atomicity transactions which are considered crucial for an e-commerce environment (Poggi et al., 2012). Database queries provides e-commerce with the scalability that is required to analyze the data and make decisions concerning the sale channels that need to be invested on the basis of the past performance (Chiu et al., 2014). Customers who make purchases through different e-commerce websites need to able to submit payment information, select product or service they desire to purchase. In parallel, the vendors should be capable of tracking the preferences and inquiries of their customers and process their orders accordingly (Grandon et al., 2011). Therefore, a well organized database query system is required for the development and maintenance of e-commerce website. In case when the web page is static, the content is displayed while the page is being created. However, every time the customer accesses a webpage, the same information is displayed on the static page (Xiao & Benbasat, 2011).

The dynamic web pages which derive little all their content from databases and data files are termed as data based web pages (Lu et al., 2010). These types of web pages are requested when the user press submit button or clicks a hyper link present on the web page form. In certain situations, static query is performed by the programs; for instance, display all items from the inventory. Even though no input is required by the user for this query, the outcomes fluctuate on the basis of the time of query (Fang, 2011). In cases, when the user clicks the submit form button instead of hyper link to make a request, then the form inputs are used by the web server program to create a query. This can be explained with the help of an example in which, the user might chose 10 books to be purchased and subsequently submits input to the web Server programs, which then processes the order and generates a dynamic web Page response to confirm the transaction (Lu et al., 2010).

Cache Early and Cache Often: Implement caching at every layer of your application. Add caching support to the data layer, the business logic layer, and the UI or output layer. Memory is cheap. Caching of data substantially reduces the load on database server. By implementing caching in an intelligent fashion throughout the application, you can achieve great performance gains. Cache can be understood with the diagram given below:



Caching used for an E-commerce Application

Using cache in any e-commerce application directly increases performance of web application.

THE TEST ENVIRONMENT

We created the test application using Visual Studio 2010 and MS-SQL Server 2008. We have used 3-tier application architecture (presentation tier, logic tier and data tier) to test the application. SQL stored procedure was used because of the cached execution plan provides stored procedures a performance advantage over normal inline SQL queries. Stored procedures are more secure. They enhance security. We create more than 67K records in the tables and tried to access the data using web application directly from the server. When we accessed the application after some waiting time we got *System Out Of Memory Exception* show below in figure 4. Same exception was generated when we tried inline SQL queries and also using stored procedure. The query used:

*SELECT * FROM PRODUCTS* and got the below exception:

Server Error in '/resrch' Application.

Exception of type 'System.OutOfMemoryException' was thrown.

Description: An unhandled exception occurred during the execution of the current web request. Please review the stack trace for more information about the error and where it originated in the code.

Exception Details: System.OutOfMemoryException: Exception of type 'System.OutOfMemoryException' was thrown.

Source Error:

[No relevant source lines]

Source File: c:\Users\Tasnim\AppData\Local\Temp\Temporary ASP.NET Files\resrch\784efb00\78dc27fd\App_Web_rpn0ou5v.0.cs **Line:** 0

Figure 4. Memory exception error

The sample stored procedure which worked fine with the application is:

```
CREATE PROCEDURE [dbo].[usp_ProductsListByName]
(@name varchar(20))
AS
BEGIN
SELECT * FROM PRODUCTS WHERE ProdName LIKE '%' + @name + '%'
END
```

Before running the web application we checked the query retrieval time at T-SQL and we got all the 67216 rows in just 1 second as shown in below figure 5:

ProdId	VendorProdId	ProdName	ProdDesc	SupplierId	CategoryId	UnitPrice	AvailableSize	AvailableColors	Size	Discount	UnitsInStock	UnitsOnOrder	ReorderLevel	ProdAvailable	DiscountAvailab
1	1	150	L. G. Mobile	L. G. Smart Mobile Phone	20	1	4500	500	black, blue, white	500	2%	450	500	100	1
2	2	150	L. G. Mobile	L. G. Smart Mobile Phone	20	1	4500	500	black, blue, white	500	2%	450	500	100	1
3	3	150	L. G. Mobile	L. G. Smart Mobile Phone	20	1	4500	500	black, blue, white	500	2%	450	500	100	1
4	4	150	L. G. Mobile	L. G. Smart Mobile Phone	20	1	4500	500	black, blue, white	500	2%	450	500	100	1
5	5	150	L. G. Mobile	L. G. Smart Mobile Phone	20	1	4500	500	black, blue, white	500	2%	450	500	100	1
6	6	150	L. G. Mobile	L. G. Smart Mobile Phone	20	1	4500	500	black, blue, white	500	2%	450	500	100	1
7	7	150	L. G. Mobile	L. G. Smart Mobile Phone	20	1	4500	500	black, blue, white	500	2%	450	500	100	1
8	8	150	L. G. Mobile	L. G. Smart Mobile Phone	20	1	4500	500	black, blue, white	500	2%	450	500	100	1
9	9	150	L. G. Mobile	L. G. Smart Mobile Phone	20	1	4500	500	black, blue, white	500	2%	450	500	100	1
10	10	150	L. G. Mobile	L. G. Smart Mobile Phone	20	1	4500	500	black, blue, white	500	2%	450	500	100	1
11	11	150	L. G. Mobile	L. G. Smart Mobile Phone	20	1	4500	500	black, blue, white	500	2%	450	500	100	1
12	12	150	L. G. Mobile	L. G. Smart Mobile Phone	20	1	4500	500	black, blue, white	500	2%	450	500	100	1
13	13	150	L. G. Mobile	L. G. Smart Mobile Phone	20	1	4500	500	black, blue, white	500	2%	450	500	100	1
14	14	150	L. G. Mobile	L. G. Smart Mobile Phone	20	1	4500	500	black, blue, white	500	2%	450	500	100	1
15	15	150	L. G. Mobile	L. G. Smart Mobile Phone	20	1	4500	500	black, blue, white	500	2%	450	500	100	1
16	16	150	L. G. Mobile	L. G. Smart Mobile Phone	20	1	4500	500	black, blue, white	500	2%	450	500	100	1
17	17	150	L. G. Mobile	L. G. Smart Mobile Phone	20	1	4500	500	black, blue, white	500	2%	450	500	100	1
18	18	150	L. G. Mobile	L. G. Smart Mobile Phone	20	1	4500	500	black, blue, white	500	2%	450	500	100	1
19	19	150	L. G. Mobile	L. G. Smart Mobile Phone	20	1	4500	500	black, blue, white	500	2%	450	500	100	1
20	20	150	L. G. Mobile	L. G. Smart Mobile Phone	20	1	4500	500	black, blue, white	500	2%	450	500	100	1
21	21	150	L. G. Mobile	L. G. Smart Mobile Phone	20	1	4500	500	black, blue, white	500	2%	450	500	100	1

Figure 5. T-SQL query completion time

After optimizing the query used in stored procedure we tried the application for different brands and products and we got the application running successful. For iPhone brand we got the search result of 16000 rows just less than 0.3 seconds as shown in the figure 6 below. The timing was only 0.28301 m. s.

Total PRODUCTS: 16000 Total Time Elapsed: 0.283016099998349

ProdId	ProdName	ProdDesc	UnitPrice	AvailableColors	Size	UnitsInStock	UnitsOnOrder	ReorderLevel	ProdAvailable	DiscountAvailable	CurrentOrder	Ranking	Note
402	iPhone 6 - 64GB	Apple iPhone 6 - 64GB	55200	black, white	500	350	400	100	True	True	50	*****	Good economical phone
403	iPhone 6 - 64GB	Apple iPhone 6 - 64GB	55200	black, white	500	350	400	100	True	True	50	*****	Good economical phone
404	iPhone 6 - 64GB	Apple iPhone 6 - 64GB	55200	black, white	500	350	400	100	True	True	50	*****	Good economical phone
405	iPhone 6 - 64GB	Apple iPhone 6 - 64GB	55200	black, white	500	350	400	100	True	True	50	*****	Good economical phone
406	iPhone 6 - 64GB	Apple iPhone 6 - 64GB	55200	black, white	500	350	400	100	True	True	50	*****	Good economical phone
407	iPhone 6 - 64GB	Apple iPhone 6 - 64GB	55200	black, white	500	350	400	100	True	True	50	*****	Good economical phone
408	iPhone 6 - 64GB	Apple iPhone 6 - 64GB	55200	black, white	500	350	400	100	True	True	50	*****	Good economical phone
409	iPhone 6 - 64GB	Apple iPhone 6 - 64GB	55200	black, white	500	350	400	100	True	True	50	*****	Good economical phone

Figure 6. SQL Query output for specific product

VI. CONCLUSION

Coming from all discussions, we can easily conclude that integration of database queries in e-commerce product searches is one of the biggest initiatives towards integrated, constant, and organized e-commerce activities. Nowadays, majority of the people worldwide prefer to use e-commerce sites for purchasing and selling products. These sites are gaining popularity among the users, as they enable to search for the desired products with ease (Geerts & Poggi, 2010). At present, the integration of optimized database queries with the e-commerce product searches is performed either manually or semi automatically (Hinz et al., 2011). The researches being performed with the proliferation of e-commerce, only describe the function and attributes of e-commerce. However, little is known about the benefits of integrating database queries with the e-commerce product searches. Nevertheless, the growing demand of e-commerce raises the need to integrate database queries in its products searches (Hinz et al., 2011). This method makes the search simpler and accurate. The principal purpose of this study was to examine how database queries can be successfully integrated with e-commerce searches. The study also analyzed the core benefits of using e-commerce websites and optimized database queries simultaneously. This research would be valuable for identifying the critical success factors for e-commerce which in turn might assist successful implementation of e-commerce sites in developing countries.

ACKNOWLEDGEMENT

We would like to gratefully and sincerely thank to The Dean of our College, and Chairman of our Departments for their guidance, understanding, patience, and most importantly, their friendly nature during this research paper writing at College of computers and Information Technology, Taif University, Saudi Arabia. We would also like to thank all of the members of the research group and friends who have always supported us for giving the opportunity to write this research paper.

REFERENCES

- [1]. R. Agrawal, A. Somani, and Y. Xu, "Storage and querying of e-commerce data", VLDB Endowment Vol. 1, pp. 149-158, September 2010. Available: <http://www.vldb.org/conf/2001/P149.pdf>, accessed Dec. 2015.
- [2]. G. I. Casterella, and L. Vijayasarathy, "An Experimental Investigation of Complexity in Database Query Formulation Tasks", Journal of Information Systems Education, Vol. 24, no. 3, Fall 2013.
- [3]. C. M. Chiu, E. T. Wang, Y. H. Fang, and H. Y. Huang, "Understanding customers' repeat purchase intentions in B2C e-commerce: the roles of

- utilitarian value, hedonic value and perceived risk”, *Information Systems Journal*, Vol. 24, no. 1, pp. 85-114, 2014. Available: <http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2575.2012.00407.x/full>, accessed Sep. 2015.
- [4]. A. Das-Sarma, N. Parikh, and N. Sundaresan, “E-commerce product search: personalization, diversification, and beyond”, In *Proceedings of the companion publication of the 23rd international conference on World wide web companion*, International World Wide Web Conferences Steering Committee, 2014. Available: <http://wwwconference.org/proceedings/www2014/companion/p189.pdf>, accessed Sep. 2015.
- [5]. S. Endrullis, A. Thor, and E. Rahm, “Entity search strategies for mashup applications”, In *Data Engineering (ICDE)*, IEEE, 2012. Available: http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6228073&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D6228073, accessed Nov. 2015.
- [6]. Y. Fang, “A model for E-commerce risk assessment with uncertain linguistic information”, *AISS: Advances in Information Sciences and Service Sciences*, Vol. 3, no. 7, pp. 296-301, Aug. 2011. Available: http://www.aicit.org/aiss/ppl/0_AISS_Binder_08_35.pdf, accessed Nov. 2014.
- [7]. F. Geerts and A. Poggi, “On database query languages for k-relations”, *Journal of Applied Logic (ELSEVIER)*, Vol. 8, no. 2, pp. 173-185, 2010.
- [8]. E. E. Grandon, S. A. Nasco, and P. P. Mykytyn Jr, “Comparing theories to explain e-commerce adoption”, *Journal of Business Research*, Vol. 64, no. 3, pp. 292-298, 2011.
- [9]. O. Hinz, I. H. Hann, and M. Spann, “Price discrimination in e-commerce? An examination of dynamic pricing in name-your-own price markets”, *MIS quarterly*, Vol. 35, no. 1, pp. 81-98, 2011. Available: http://www.ecm.bwl.uni-muenchen.de/publikationen/pdf/price_discr.pdf, accessed Nov. 2015.
- [10]. S. Li and E. Karahanna, “Peer-based recommendations in online B2C e-commerce: comparing collaborative personalization and social network-based personalization”, *IEEE*, Jan. 2012. Available: http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6148983&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D6148983, accessed Oct. 2015.
- [11]. Y. Lu, L. Zhao, and B. Wang, “From virtual community members to C2C e-commerce buyers: Trust in virtual communities and its effect on consumers’ purchase intention”, *Electronic Commerce Research and Applications*, Vol. 9, no. 4, pp. 346-360, Aug. 2010. Available: <http://aeo4me.com/2012/carey/%E6%96%87%E7%8D%BB/0924/Trust%20in%20virtual%20communities%20and%20its%20effect%20on%20consumers%E2%80%99%20purchase%20intention.pdf>, accessed Jun. 2015.
- [12]. N. Poggi, D. Carrera, R. Gavalda, E. Ayguadé, and J. Torres, “A methodology for the evaluation of high response time on E-commerce users and sales”, *Information Systems Frontiers*, pp. 1-19, Oct. 2012. Available: <http://link.springer.com/article/10.1007/s10796-012-9387-4#page-1>, accessed Jan. 2016.
- [13]. D. Vander Meer, K. Dutta, and A. Datta, “A cost-based database request distribution technique for online e-commerce applications”, *MIS quarterly*, Vol. 36, issue 2, pp. 479-508, Jun. 2012.
- [14]. B. Xiao and I. Benbasat, “Product-related deception in e-commerce: a theoretical perspective”, *MIS Quarterly*, Vol. 35, issue 1, pp. 169-196, Mar. 2011. Available: <http://dl.acm.org/citation.cfm?id=2017493>, accessed Dec. 2015.
- [15]. E. F. Codd, “A Relational Model of Data for Large Shared Data Banks”. Available at: <https://www.seas.upenn.edu/~zives/03f/cis550/codd.pdf>, accessed March 2016

Evaluation of Machine Vision for robot navigator: mini review

Arefe Esalat Nejad

Young Researchers and Elite Club, Baft Branch, Islamic Azad University, Baft, Iran.

*Corresponding Author

Abstract

Machine vision (MV) is the technology and methods used to provide imaging-based automatic inspection and analysis for such applications as automatic inspection, process control, and robot guidance in industry. This paper presents some of the underlying concepts and principles that were key to the design of our research robots. Vision is an ideal sensor modality for intelligent robots. It provides rich information on the environment as required for recognizing objects and understanding situations in real time. Moreover, vision-guided robots may be largely calibration-free, which is a great practical advantage. Three vision-guided robots and their design concepts are introduced: an autonomous indoor vehicle, a calibration free manipulator arm, and a humanoid service robot with an omnidirectional wheel base and two arms. Results obtained, and insights gained, in real-world experiments with them are presented. Researchers and developers can take it as a background information for their future works.

Key words: Machine vision (MV), Intelligence robots, human service, Robot guidance

Introduction

Since the end of the 18th century with the first Industrial Revolution through the introduction of mechanical production facilities powered by water and steam, factories have experimented big changes in their production systems [1]. The second Industrial Revolution, in the start of the 20th Century, introduced mass production based on the division of labor powered by electrical energy [2]. The third Industrial Revolution of the start of 1970s introduced the use of electronics and information

technologies for a further automatization of production [3]. Nowadays, we are involved in the fourth Industrial Revolution, commonly called “Industry 4.0”, based on cyber-physical production systems (CPS) and embracing automation, data exchange and manufacturing technologies. These cyber-physical systems monitor the physical processes, make decentralized decisions and trigger actions, communicating and cooperating with each other and with humans in real time. This facilitates fundamental improvements to the industrial processes involved in manufacturing, engineering, material usage and supply chain and life cycle management [4].

Although present robots contribute very much to the prosperity of the industrialized countries they are quite different from the robots that researchers have in mind when they talk about “intelligent robots”. Today’s robots

- <are not creative or innovative,
- <do not think independently,
- <do not make complicated decisions,
- <do not learn from mistakes,
- < do not adapt quickly to changes in their surroundings.

They rely on detailed teaching and programming and carefully prepared environments. It is costly to maintain them and it is difficult to adapt their programming to slightly change environmental conditions or modified tasks. Although the vast majority of robots today are used in factories, advances in technology are enabling robots to automate many tasks in non-manufacturing industries such as agriculture, construction, health care, retailing and other services. These so-called “field and service robots” aim at the fast growing service sector and promise to be a key product for the next decades. From a technical point of view service robots are intermediate steps towards a much higher goal: “personal robots” that will be as indispensable and ubiquitous as personal computers today. Personal robots must operate in varying and unstructured environments without needing maintenance or programming. They must cooperate and coexist with humans who are not trained to cooperate with robots and who are not necessarily interested in them. Advanced safety concepts will be as indispensable as intelligent communication abilities, learning capabilities, and reliability. It will be a long way

of research to achieve this goal, but undoubtedly vision – the most powerful sensor modality known – will enable these robots to perceive their environments, to understand complex situations and to behave intelligently. This paper presents some of the underlying concepts and principles that were key to the design of our research robots.

In brief, they are:

<Vision is the most powerful sensor modality for providing rich and timely information on a robot's environment.

<Behavior is the key to a powerful system architecture that enables a robot to construct complex actions by combining elementary behavior primitives.

<Situation assessment is the basis for the dynamic selection of the most appropriate behavior by a robot in its interactions with the outside world.

<Perception rather than measurement should be the basis for situation assessment and robot control.

We expect that these fundamental concepts are a strong basis for future generations of intelligent robots that combine locomotive and manipulative actions. In section 2 these concepts are explained in more detail. Another fundamental principle has considerably influenced our research work: Every result has to be proved and demonstrated in practical experiments and in the real world.

Vision and its Potential for Robots

When a human drives a vehicle he depends mostly on his eyes for perceiving the environment. He uses his sense of vision not only for locating the path to be traversed and for judging its condition, but also for detecting and classifying external objects, such as other vehicles or obstacles, and for estimating their state of motion. Entire situations may thus be recognized, and expectations, as to their further development in the “foreseeable” future, may be formed. The same is true for almost all animals. With the exception of those species adapted to living in very dark environments, they use vision as the main sensing modality for controlling their motions. Observing

animals, for instance, when they are pursuing prey or trying to escape a predator, may give an impression of the performance of organic vision systems for motion control. In some modern factory and office buildings mobile robots are operating, but almost all of them are blind. Their sensors are far from adequate for supplying all the information necessary for understanding a situation. Some of them have only magnetic or simple optical sensors, allowing them merely to follow an appropriately marked track. They will fail whenever they encounter an obstacle and they are typically unable to recover from a condition of having lost their track. The lack of adequate sensory information is an important cause making these robots move in a comparatively clumsy way and restricting their operation to the simplest of situations. Other mobile robots are equipped with sonar systems. Sonar can, in principle, be a basis for powerful sensing systems, as evidenced by certain animals, such as bats or dolphins. But the sonar systems used for mobile robots are usually rather simple ones, their simplicity and low cost being the very reason for choosing sonar as a sensing modality. It is then not surprising that such systems are severely limited in their performance by low resolution, specular reflections, insufficient dynamic range, and other effects [7].

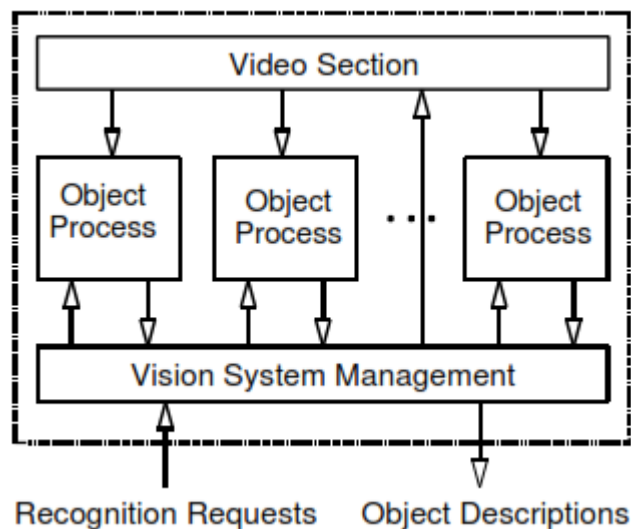


Figure 1: Conceptual structure of object-oriented robot vision systems.

Likewise, it may be expected that advanced robots of the future will also rely primarily on vision for perceiving their environment, unless they are intended to operate in other environments, e.g. under water, where vision is not feasible.

One apparent difficulty in implementing vision as a sensor modality for robots is the huge amount of data generated by a video camera: about 10 million pixels per second, depending on the video system used. Nevertheless, it has been shown (e.g., by [5]) that modest computational resources are sufficient for realizing real-time vision systems if a suitable system architecture is implemented. As a key idea for the design of efficient robot vision systems the concept of object-oriented vision was proposed. It is based on the observation that both the knowledge representation and the data fusion processes in a vision system may be structured according to the visible and relevant external objects in the environment of the robot (Figure1).

Behavior

Biological behaviors could be defined as anything that an organism does involving action and response to stimulation, or as the response of an individual, group, or species to its environment. Behavior-based robotics has become a very popular field in robotics research because biology proves that even the simplest creatures are capable of intelligent behavior: They survive in the real world and compete or cooperate successfully with other beings. Why should it not be possible to endow robots with such an intelligence? By studying animal behavior, particularly their underlying neuroscientific, psychological and ethological concepts, robotic researchers have been enabled to build intelligent behavior-based robots according to the following principles:

- <complex behaviors are combinations of simple ones, complex actions emerge from interacting with the real world
- <behaviors are selected by arbitration or fusion mechanisms from a repertoire of (competing) behaviors
- <behaviors should be tuned to fit the requirements of a particular environment and task
- < Perception should be actively controlled according to the actual situation

Situation Assessment

According to the classical approach, robot control is model-based. Numerical models of the kinematics and dynamics of the robot and of the external objects that the robot should interact with, as well as quantitative sensor models, are the basis for controlling the robot's motions. The main advantage of model-based control is that it lends itself to the application of classical control theory and, thus, may be considered a straight-forward approach. The weak point of the approach is that it breaks down when there is no accurate quantitative agreement between reality and the models.

Figure 2 illustrates the definition of the term “situation” by embedding it in the action-perception loop of a situation oriented behavior-based robot. The actions of the robot change the state of the environment, and some of these changes are perceived by the robot's sensors. After assessing the situation an appropriate behavior is selected and executed, thus closing the loop. The role of a human operator is to define external goals via a man machine interface and to control behavior selection, e.g., during supervised learning.

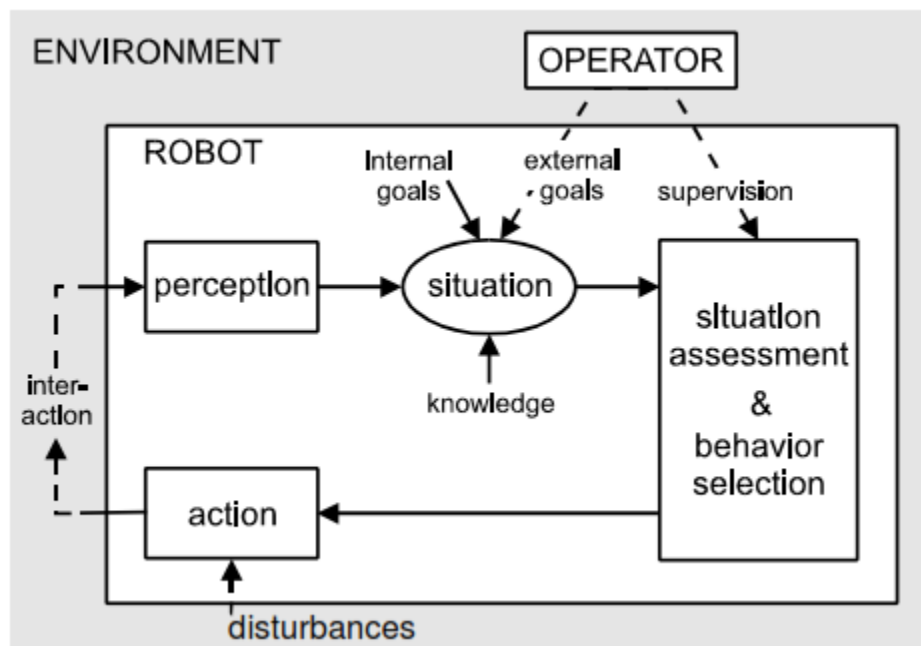


Figure 2: The role of “situation” as a key concept in the perception-action loop of a situation-oriented behavior-based robot

Realized System Architecture

Figure 3 gives an overview of the system architecture that has been realized for our situation-oriented behavior-based robot. In this section we give only a short introduction to the different modules and their interaction (see [6] for details):

<a situation module taking into account all the decisive factors explained in section 2.3 and basing thereupon the dynamic selection of behaviors

<a sensor module comprising an object-oriented vision system as the main sensor and a proprioceptor system that provides auxiliary information needed by certain behavior patterns

< An actuator module executing commanded behaviors by activating a sequence of control laws for the drives

<an extendable knowledge base providing information about the static characteristics of the environment and the actual mission and goals

< A man machine interface for operator intervention and status display

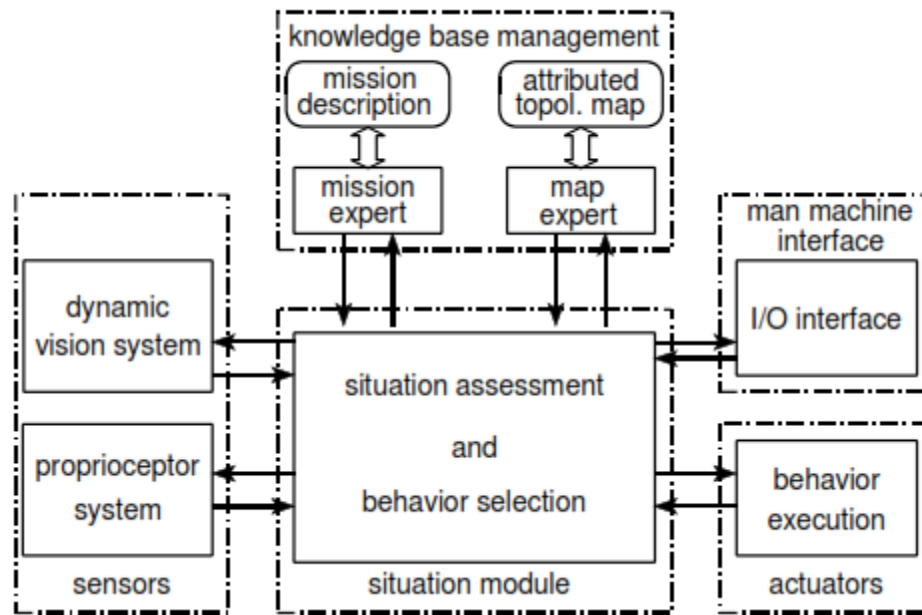


Figure 3: System architecture of a situation-oriented behavior-based robot.

Conclusions

In this paper our fundamental concepts and principles for designing and building intelligent robots have been presented. We strongly believe that vision – the sensor modality that predominates in nature – is also an eminently useful and practical sensor modality for robots. It provides rich and timely information on the environment and allows real-time recognition of dynamically changing situations. Situation-dependent perception and behavior selection rather than measurement and control based on quantitatively correct models are additional key factors for advanced robots. Motor control commands should be derived directly from sensor data, without using world coordinates or parameter-dependent computations, such as inverse perspective or kinematic transforms.

References

1. Deane, P.M. The First Industrial Revolution; Cambridge University Press: Cambridge, UK, 1979.

2. Kanji, G.K. Total quality management: the second industrial revolution. *Total Qual. Manag. Bus. Excell.* 1990, 1, 3–12.
3. Rifkin, J. The third industrial revolution. *Eng. Technol.* 2008, 3, 26–27.
4. Kagermann, H.; Wahlster, W.; Helbig, J. Recommendations for Implementing the Strategic Initiative Industry 4.0: Final Report of the Industry 4.0 Working Group; Forschungsunion: Berlin, Germany, 2013.
5. Graefe, V. Dynamic Vision Systems for Autonomous Mobile Robots. *Proc. IEEE/RSJ International Workshop on Intelligent Robots and Systems, IROS '89.* Tsukuba, 1989, pp. 12-23.
6. Bischoff, R.; Graefe, V.; Wershofen, K. P. Object-Oriented Vision for a Behavior-Based Robot. In D. Casasent (ed.): *Intelligent Robots and Computer Vision XV.* Boston, November 1996, *Proc. SPIE*, Vol. 2904, 1996, pp. 278-289.
7. Koeppe, R. New industrial robotics: human and robot collaboration for the factory. In *Proceedings of the 2014 European Conference on Leading Enabling Technologies for Societal Challenges (LET'S 2014)*, Bologna, Italy, 29 September–1 October 2014.

Significant Approach for Detaining Unpromising Contestant for Mining Pattern Based on Profit

Vijay Kumar Verma

Lord Krishna College of Technology
Department of Computer Science and Engineering.
Indore M.P. India

Kanak Saxena

Samrat Ashok Technological Institute
Department of Computer Applications
Vidisha M.P. India

Abstract: -Today's every business organization needs profit. Professionals might give attention on recognizing its most treasured consumers. Consumers who give a major portion of the profits to the business. Frequency based mining of items do not fulfill all the requirements of business. They only provide the information that an item has high low frequency based on a given value. There is one important factor profit has to be considered by every business. In past years a lot of methods have been developed for mining profit based patterns but efficiency, accuracy and scalability are important factors that have always to be considered. In this paper we proposed a significant approach for detaining unpromising contestants for mining profit based patterns. The proposed approach mines profit based patterns accurately and removes all unpromising contestants at different levels.

Keywords :- Profit, Pattern, unpromising, frequency, efficiency

I. INTRODUCTIONS

Online purchasing is a common habit today. Real life applications every day generate huge amounts of data and discovering important information from this data is a difficult task. Data mining provides several methods and techniques to mine meaningful information from huge amounts of data. Frequency based pattern mining is one of the important techniques for mining patterns based on frequency as per the given support threshold. Frequency based patterns do not fulfill all the requirements from the business point of view. An item has several dimensions like profit, cost, time and quantity, so considering this parameter is also an important issue.

II. BASIC CONCEPTS

Profit mining of an item set is the discovery of finding all those items from the dataset which has profit more than given threshold. Profit mining contains several interrelated terms. In this section we define these terms one by one.

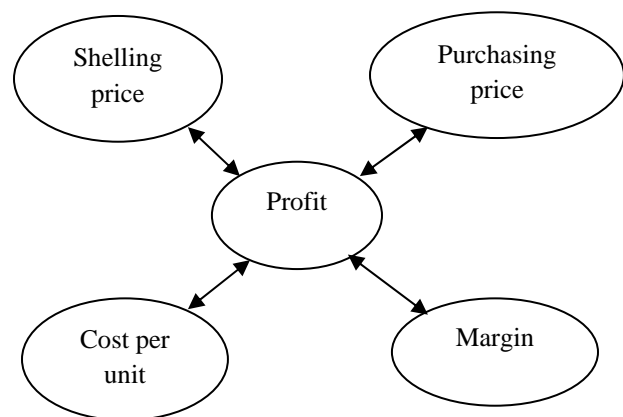


Figure 1. Meaning of profit

Let E is a set of elements denoted by $E = \{e_1, e_2, e_3, \dots, e_n\}$.

Let all purchasing records be denoted by CPR

$CPR = \{R_1, R_2, R_3, \dots, R_n\}$ where each record $R_i \in CPR$.

DB is set of all records of customer who are purchasing these elements. Each element has a profit value denoted by p .

TABLE 1 PURCHASING RECORD

TID & Element	e_1	e_2	e_3	e_4	e_5
R1	0	0	18	0	1
R2	0	6	0	1	1
R3	2	0	1	0	1
R4	1	0	0	1	1
R5	0	0	4	0	2
R6	1	1	0	0	0
R7	0	10	0	1	1
R8	3	0	25	3	1
R9	1	1	0	0	0

R10	0	6	2	0	2
-----	---	---	---	---	---

TABLE 2 PROFIT TABLE

ITEM	PROFIT(\$)
e ₁	3
e ₂	10
e ₃	1
e ₄	6
e ₅	5

III. RELATED TERMINOLOGY

The objective of profit based pattern is to find out all those elements or set of elements which has the profit values beyond a user specified threshold in database. There are several terms are related with profit based pattern, in this sections we introduce these term with notations.

Definitions 1.

Local profit of an element represented by a numeric value e_{ip} in purchasing records R_q . For example, $lp(e_1, R_8) = 3$, in table 1.

Definitions 2.

External profit of an element is independent of records and associated with item e_{ip} in the profit table. This value shows the importance of an element. For example, in table 2 the external profit of item e_1 , $ep(e_1)$, is 3.

Definitions 3.

The profit of an element e_{ip} in a purchasing record R_i is the numerical measure and calculated by with profit function denoted by $p(e_{ip}, R)$. For example profit of item e_5 in record R_5 is $2 * 5 = 10$.

Definitions 4.

The profit of an element set S in record R is calculated by multiplying quantity of element with profit. For example profit of element set $\{e_2, e_5\}$ in record R_2 is defined as $pes(\{e_2, e_5\}, R_2) = p(\{e_2\}, R_2) + p(\{e_5\}, R_2) = 6 * 10 + 1 * 5 = 65$.

Definitions 5.

The profit of element set S in database is the numerical measure and calculated by sum of all value of records where the element presents. For example the profit of an element set element $\{e_1, e_5\}$ in database $pesd(\{e_1, e_5\}) = u(\{e_1, e_5\}, R_3) + u(\{e_1, e_5\}, R_4) + u(\{e_1, e_5\}, R_8) = 33$.

Definitions 6.

The profit of purchasing record R is calculated by the sum all element's profit in that record. For example profit of record $R10$ is defined as $aspr(R10) = p(\{e_2\}, R10) + p(\{e_3\}, R10) + p(\{e_5\}, R10) = 72$.

Definitions 7.

The profit of entire database is calculated by sum of all records profit. For example the profit of database is $p(CPR) = p(R1) + \dots + P(R10) = 23 + \dots + 72 = 400$.

IV. LITERATURE REVIEW

In the past year several researcher have been proposed various method for mining profit based pattern. The objective of each and every method is to increase efficiency. We study some of paper related of our proposed works.

In 2005 Hong, Hamilton, and Cory first time introduced the concepts of the profit based pattern. They give a theoretical model for mining profit based pattern. They introduced the concepts support bound property concepts. They give a mathematical model for mining profit based pattern.

In 2005 Liu and Liao introduce Two-Phase method to mine profit based pattern. They used the concepts of records based profit. They introduce a model that is based on downward closure property.

In 2007 Erwin and Gopalan introduced Bottom-Up Projection Based approach for mining profit based pattern. They create a structure and call compressed profit pattern. They also used GlobalItemTable for string the item, index, profit and quantity.

In 2008 Erwin Gopalan and Achuthan proposed that anti-monotone property leads to a larger search space and there for compact utility pattern tree data structure are needed. They introduce TWU based pattern growth tree for mining profit based pattern. They also used parallel projection scheme.

In 2009 Chowdhury, Syed and Jeong proposed HUC-Prune approach for mining high profit based pattern. They used tree-based candidate pruning technique. They also used a hash table and properties of FP tree in proposed method.

In 2010 Tseng, Bai-En and Philip S. proposed UP-Growth tree based approach for mining profit based pattern. They proposed IHUP Tree Structure. They also used lexicographic order to rearranged transactions. They used a table which store profit and link of nodes.

In 2011 S. Kannimuthu Premalatha and Shankar proposed iFUM. They improve the FUM approach by applying combination generator functions. This function generates the combination of a set so that all its sub set are also satisfy the condition for profit based pattern.

In 2012 Srinivasa Rao Krishna Prasad Improved UP-Growth approaches for mining profit based pattern. They used a global UP-Tree for discarding global unpromising items. They apply DGNstrategy for the node utilities which are nearer to UP-Tree root node are effectively reduced.

In 2013 Arumugam P and Jose P proposed advanced concepts for mining high profit pattern using electronics data set. They used Transaction-weighted Downward Closure Property in proposed approach.

In 2014 Philippe Cheng and Vincent proposed Co-occurrence based Pruning strategy for mining profit based pattern. They introduced co-occurrence to reduce the number of join operations that need to be performed. EUCP (Estimated Utility Co-occurrence Pruning) is based on the observation that most costly operation in HUI-Miner is the join operation.

V. PROBLEM STATEMENT

After study some of the paper related to our topic we found that there are three main problems that has to consider for improving the performance of the existing algorithms.

A. Unpromising elements

Discarding unpromising element is a crucial issue. Useless element increases the search apace and need more join operations.

B. Efficiency

How accurately profit based pattern has to be generated is a big issue because most of the algorithms are based on records based profit but element has self-utility which are not consider.

C. Complexity

How Complexity can be reduced is also a difficult task. Complexity may arithmetic or space complexity reducing these complexity increase the performance of the algorithms

VI. PROPOSED APPROACH

Our proposed approach is based on two important factors first factors is reducing search space and seconds reducing size of records.

1. Decreasing search space

We know that if we have n element then we need to check there are 2^n sub set. Similarly if we have k unpromising candidate then we need to also check $2^{(n-k)}$ uses less subset. Our objective is remove the unpromising candidates at initial level so that there is no need to create pairing and also no need check there subset.

For example if we have three items {a, b, c}

Then the number of sub set is

{{ }, {a}, {b}, {c}, {a, b}, {a, c}, {b, c}, {a, b, c}}.

Now suppose a is unpromising candidates then we need to check $2^{(n-k)}$ sub set which is useless. Here k=1 and n=3

$$2^{(n-k)} = 2^{(3-1)} = 4$$

{{ }, {a}, {b}, {c}, {a, b}, {a, c}, {b, c}, {a, b, c}}.

2. Reducing number of records

Purchasing records contain element. To reduce the searching time, the records with the same items are grouped into single records and the number of element in the grouped records is also added. This process reduces calculation time, reduces search space and improves execution time.

VII. PROPOSED ALGORITHM

The proposed algorithm has been divided into the following steps

Step1. Calculate the profit value of each element.

Step2. Calculate the profit of each purchasing record.

Step3. Calculate the total profit of all records.

Step4. Calculate the profit of each record for each element.

Step5. Calculate the self-profit of each element the sum of the profit values of the records

Step6. Now check the records based profit of an element and self-profit, if it is greater than or equal with given value then high profit element, otherwise not high profit element.

Step7. After first find unpromising element and delete from the database. Now group the records with the similar elements.

Step 8 Repeat this process until no more item remains in the database

Step9. Exit.

VIII. EXPERIMENTAL ANALYSIS

We have implemented proposed algorithm with TP and iFUM algorithms using C3 dot net 2010. We used 25 different item and 1000 records form an electronic shop. We SQL server 2010 R2 for storing the database. We used real life dataset. We windows 7 operating system with i3 processor.

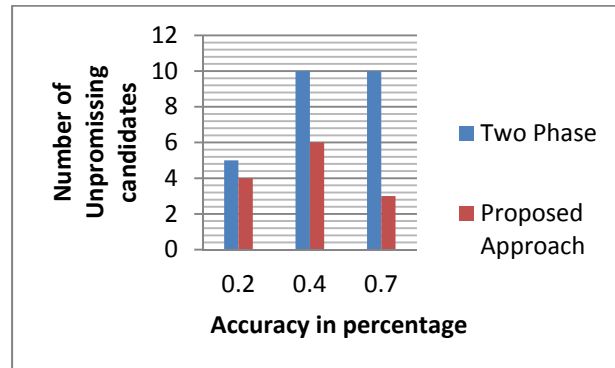


Figure 1 accuracy in percentage

REFERENCE

- [1] 2005 Hong Yao, Howard J. Hamilton, and Cory J. Butz "A Foundational Approach to Mining Itemset Utilities from Databases" Department of Computer Science University of Regina Canada.
- [2] Ying Liu Wei-keng Liao Alok Choudhary "A Fast High Utility Itemsets Mining Algorithm" UBDM, August 21, 2005, Chicago Copyright 2005 ACM 1-59593
- [3] Alva Erwin, Raj P. Gopalan, N.R. Achuthan "A Bottom-Up Projection Based Algorithm for Mining High Utility Itemsets" 2007, Australian Computer Society, Inc.
- [4] Alva Erwin, Raj P. Gopalan, and N.R. Achuthan "Efficient Mining of High Utility Itemsets from Large Datasets PAKDD 2008, LNAI 5012, pp. 554-561, 2008. Springer-Verlag Berlin Heidelberg 2008
- [5] Chowdhury Farhan Ahmed, Syed Khairuzzaman Tanbeer, "An Efficient Candidate Pruning Technique for High Utility Pattern Mining" T. Theeramunkong PAKDD 2009, LNAI 5476, 2009. Springer-Verlag Berlin Heidelberg 2009
- [6] Vincent S. Tseng, Cheng-Wei Wu, Bai-En Shie, and Philip S. Yu "UP-Growth: An Efficient Algorithm for High Utility Itemset Mining" KDD'10, July 25-28, 2010, Washington, DC, USA. Copyright 2010 ACM 978-1-4503-0055-1/10/07
- [7] S. Kannimathu Dr. K. Premalatha iFUM - Improved Fast Utility Mining International Journal of Computer Applications (0975 - 8887) Volume 27- No.11, August 2011
- [8] Adinarayanareddy B. O.Srinivasa Rao, PhD An Improved UP-Growth High Utility Itemset Mining International Journal of Computer Applications (0975 - 8887) Volume 58- No.2, November 2012
- [9] Arumugam P and Jose P "ADVANCE MINING OF HIGH UTILITY ITEMSETS IN TRANSACTIONAL DATA "Opinion - International Journal of Business Management (e-ISSN: 2277-4637 and p-ISSN: 2231-5470) Special Issue on Role of Statistics

in Management and Allied Sciences Vol. 3 No. 2 Dec. 2013, pg. 27-40

- [10] More Rani N. Anbhule Reshma V "Mining High Utility Item sets From Transaction Database" International Journal of Latest Trends in Engineering and Technology (IJLTET) Issue 3 January 2014 180 ISSN: 2278-621X
- [11] Philippe Fournier-Viger, Cheng-Wei Wu, Souleymane Zida, Vincent S. Tseng" (2014) FHM: Faster High-Utility Itemset Mining using Estimated Utility Co-occurrence Pruning. Proc. 21st International Symposium on Methodologies for Intelligent Systems (ISMIS 2014), Springer, LNAI, pp. 83-92.

Scalable and Secure Network Storage in Cloud Computing

Muhib Ahmad Khan, M. Munwar Iqbal, Fahad Ubaid, Rashid Amin, Asima Ismail
Department of Computer Science, University of Engineering and Technology Taxila Pakistan

ABSTRACT

Cloud Computing is a newly born type of computation, which depends on the shared resources of the network. Cloud Computing term discovered from that time when the system can access the different types of applications as well as different types of services remotely. Cloud Computing is the unique, next generation of IT architecture, in which computation is done on the open network shared resources, which create a security risk. In comparison to the existing conventional infrastructure, The IT services come under the IT expert control. In a market there is a different type of service provider using cloud computing features offers many different services like virtualization, applications, servers, data sharing, and try to the reduce client-side computation overhead. Nevertheless, most of these services are outsourced to the third party, which creates the risk of data confidentiality as well as the data integrity. These days cloud computing, and its security is the hot topic for the research. In this paper, a new model proposed for storage data on the network for the secure data storage on the cloud server, which achieve the security, availability, confidentiality and integrity.

Keywords -- Cloud Computing, Data Integrity & Security, Data Confidentiality & Availability.

INTRODUCTION

The changing mode of technology and the rapid increase in these technologies had made the

world a global village. The emergence of the new computing technologies has with certain types of benefits as well as challenges. Cloud computing is one of unique technology which emerge with a high amount of benefits. Cloud computing comes with the combination of the other core computing technologies [5]. Cloud-based computing is more than an IT shifted standard, it converts not only the IT sector, moreover every industry of the society. In simple language, Cloud Computing is a collection and combination of different computing applications and services from different servers on a network [1] [3]. Cloud Computing is the emerging field of computer science which required more research. Due to the miraculous success of the Internet, computing resources is now more abundantly available. The term “cloud” is used as a metaphor for the internet. The basic objective of cloud computing is secure data storage and for the internet computing devices [9] [6]. In cloud computing traditional service provider follow two different ways and these are infrastructure and service provider. In infrastructure, provider arranges cloud platform and lease resources according to the demand from the service provider. In service provider take the service from the infrastructure and sale it to the end users. Cloud computing is omnipresent. Basically it comes as new era technology which gives the facility of on-demand approach to the required network. Cloud computing comes with enormous benefits which all are available on one platform such as distributed computing, virtualization and much more [9]. All the advantages which come by

gaining the cloud computing methodology, in spite of all these features the sole of the cloud is not properly easy to implement. However, Cloud is in its initial phase, faces several risks, among all the issues of the cloud computing the most risk is security [3] [4]. The old conventional data security techniques are not satisfactory. The integrity of the data cannot be achieved by the old conventional methodologies. The user transfers their critical data to the scattered cloud environment [1] [6]. Thus, the cloud provider should enforce the appropriate security protocols to protect the essential integrity, authentication and authorization protocols of data.

Although cloud computing provides many facilities in term of data storage and online computation, there are also several issues which should be handled carefully. Traditionally Security measures are not adequate enough to keep the data safe according to the data security demands. To ensure the security factor in cloud computing, we need to define more security procedure in cloud computing as compared to the recent traditional procedures.

CLOUD SECURITY

Among besides all other problems in cloud computing, the security of the data is the core issue with respect to the business model tracked by privacy, integrity, and availability. Now the security of data is main interest among various services provider organizations, especially in a shared environment.

In the public or common cloud situation, its cloud service provider responsibility ensure the adequate security protocols to the critical data regarding authentication, integrity, and compliance. There are three different types of clouds available such as [2] [3]:

1. Software as a service (SaaS)
2. Platform as a service (PaaS)
3. Infrastructure as a service (IaaS)

The other types of cloud environment which are also used as services by a huge amount of users are as under:

- A. Storage as a service
- B. Database as a service
- C. Information as a service
- D. Process as a service
- E. Integration as a service
- F. Security as a service
- G. Management/Governance as a service
- H. Testing as a service

All these three clouds based services drives on different security problems. In IaaS, the basic resources processing and network utilization are offered by the service provider where user install and run the different applications. Moreover, in IaaS, the users have a superior hold on all over the security with respect to the other models. According to the PaaS, users are able to install their software on the cloud structure without the deployment of any other additional tools and the service providers dealing in PaaS also wishes to protect the platform software stack. SaaS users use the cloud service provider software with the help of web browser. In this model, the security of data is a chief challenge when the user utilizes SaaS based model of cloud.

In cloud computing environment, the data of the users are managed, and stored as plain or simple text and backup of data are also a serious concern.

ISSUES IN CLOUD

Cloud computing is a new era computing which makes it unique. Cloud computing emergence is very rapid. As the cloud is becoming popular, it also faces some issues which make it enhanced

for the stand alone in the market. These issue of the date are related to the data level. Cloud users move their sensitive data to the cloud to make it secure, but if cloud fails to provide better security to the data, which makes the cloud improper computing. The threats related to the data are [3]:

1. Malicious Insiders
2. Denial of Service
3. Data Loss or Data Leakage
4. Data Scavenging
5. Customer-data
6. Manipulation

DATA STORAGE IN CLOUD

The Cloud Service Providers offer two basic things one is computing and second is storage [3]. In cloud computing environment data is kept at the service provider location, and it is maintained by the further distributed vendors companies. Cloud computing changes the mode of the storage of the sensitive data on the cloud where you access it remotely rather than on the Hard Disk Drives of your personal computers. The trend of storing sensitive data in the cloud, the security measure shifts from as well and it needs more security parameters than local Hard Disk Drives. Different service providers of the Cloud own huge-sized data centers for data storage. So the user whether purchase or rent some of its portion for the storage of their critical data [3] [4]. A single data center contains hundreds of thousands of servers that are arranged in a rack of 20-40 servers each. The storage providers contain the hundreds of the datacenters linked with one another form a huge structure. Data is stored and maintained in the datacenter. Storage providers provide a huge number of services; these services are bitterly use by the special Application programming interface (API) through the network. These API are specially designed for the cloud and their users. The API of the cloud provides the whole image of the cloud. It describes the cloud performance as well as its security and much more [3]. To enhance the security of the cloud,

we need to increase API efficiency and keep it update. As the cloud emerge, it proves itself more powerful, scalable, and more optimized than other available technologies [7]. Especially it has the strength or power to accommodate itself to new upcoming variations that based on the requirements, moreover the reduction of the huge amount of cost. There are different flaws in security parameters in cloud computing which make it unpopular. If it is unable to provide maximum security to the sensitive data, the all other benefits of the cloud have no value, and no one will agree to use it and make a compromise on security parameter [3].

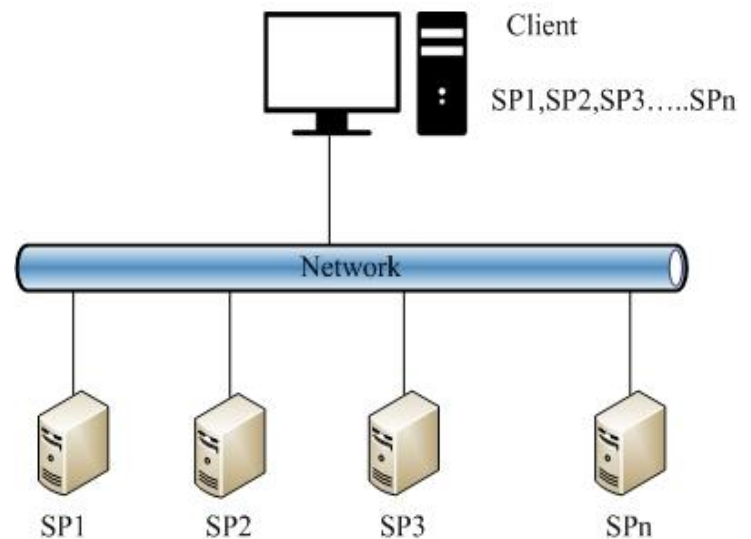


Figure 1.1: Data on different cloud services providers

The storage on the cloud offers security to the critical data by dividing the data into smaller chunks and store them in different places on the data center. If any particular chunks of the data are crashed in datacenter than remaining chunks, also sum up the data. The storing of the sensitive data as a plain text on the cloud providers location makes the data highly unsafe [3] [4] [5] [7]. The fame of storing data on the cloud is highly increasing by the reason of it can accessible from remote locations as well as it's not user concern to hold it all the time just as a reason of service provider accepts all the

accountability and all the concerns about the security.

In this paper, we suggest a more efficient way of more scalable as well as secure network storage architecture which contains different cloud storage providers for the improved storage of critical data. For better security, confidentiality, reliability and availability of the cloud, firstly the data of the user is encrypted and then divide the whole encrypted data into different chunks, then finally store it on to the different cloud services provider location. A cipher key is placed with every chunk of the data. Moreover, one of the special service providers contains whole encrypted data, but it does not contain the cipher key along with the data. If the data on any particular service provider is crashed, it can recover from this special service provider which contains all encrypted data by using the cipher key from the other service provider. For improved reliability and availability of data stored in the cloud, the Redundant Array of Inexpensive Disks model is implemented on the service provider side.

RELATED WORK

Fawaz in the related research [3] divide the data into a different number of chunks. These data distributions are places the on several different locations of clouds in such a way, that if any hacker or unauthorized person is able to gain access to a particular network. Then this unauthorized person is unable to extract the meaningful information because it is a small chunk of data and the other chunks of that critical data are stored in another different cloud locations. In other papers, the authors discussed the new advanced technology named RAID for the storage in the cloud. The Cryptographic technology is not so much mature. As the data increases in the cloud, then it is unable to provide the maximum security and privacy.

In this paper [3], the authors firstly apply the encryption mechanism on the critical data which user wishes to store in the cloud. After

encryption, the author breaks the data into cipher chunks. Chunks of the encrypted data are now placed on the different cloud provider's locations named as SP1, SP2, and SP3. Figure 1.2 shows the initial image of the author proposed architecture, containing the host machine and the different service providers [3]. The author here introduces the parity bit to restore the encrypted data. For better availability and performance of the cloud he adopt the RAID technology and implement it on every server in the datacenter for improving availability and cloud performance. Our suggested model is somehow same to this methodology with little change in the distribution strategy. According to the paper authors [3] [5] [10], RAID 10 is more efficient than other RAID models. RAID 10 provides better availability as well as performance than other models of RAID.

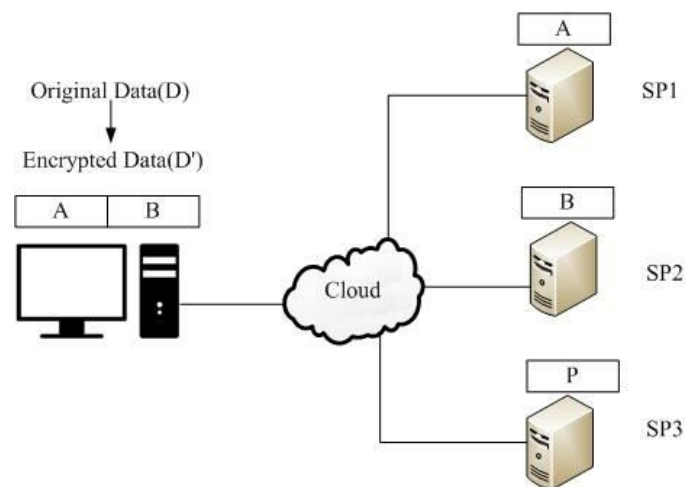


Figure 1.2: Architecture of related Model

RAID 10 (1+0) gives various characteristics like, availability, redundancy of data and fault tolerance by the combination of characteristics of the mirroring and stripping [3].

PROPOSED MODEL

In this proposed model the “scalable and secure storage in cloud computing” is achieved by the number of different steps, every step involved in this architecture has its importance. Each step involved in this architecture is designed to provide maximum security to the sensitive data

which the user want to store on the cloud and not any unauthorized person or intruder gain the access of the whole data. Firstly, the critical data is divided into a different number of chunks depending upon the length of the data. After splitting the sensitive data into chunks, an encryption mechanism is applied on all these chunks [7]. Every particular chunk has its encryption key through which the encrypted data is retrieved into its original, meaningful form. As we stated earlier that every chunk of data has its encryption key so as the number of the data chunks increases, the number of keys also increases. The conversion of the plain critical data into numbers of encrypted cipher blocks is to provide maximum security to the data. RAID 10 model is implemented at data centers that are located on the clouds services provider side to provide better availability and performance [3] [10].

Suppose D is the original data of a user which is very secret from the user opinion. The user wants to move the data to the clouds service provider location. Firstly the original data D is split into a number of blocks A, B, C, D, E, F. and G. Then, after the splitting of data D, an encryption mechanism is applied on all these blocks of the data. Which convert the data blocks to the cipher blocks A', B', C', D', E', F' and G'. Every cipher block has its encryption key K (A), K (B), K (C), K (D), K (E), K (F) and K (G). Now the cipher blocks are placed on the different cloud provider's locations named as SP1, SP2, and SP3. For a while, the chunks of the encrypted data A' is placed on SP1. B' is placed on SP2, C' is placed on SP3, D' is placed on SP4, E' is placed on SP5, F' is placed on SP6 and G' is placed on SP7. It depends on the length as well as the number of chunks of the encrypted critical data, and the key of the encrypted cipher data is

placed in such a way that a particular chunk contains the next two keys of the cipher block as shown in figure 1.3.

Original Data (D) -> Data Blocks (A, B, C, D, E, F & G) -> Encrypted Data Blocks (A', B', C', D', E', F' and G')

The figure 3 shows that the A' is on SP1 and contains the key of B. while C as K (B), K (C), B' is on SP2 and contains the key of C and D as K (C), K (D). C' is on SP3 and contains the key of D and E as K (D), K (E). D' is on SP4 and contains the key of E and F as K (E), K (F). E' is on SP5 and contains the key of F and G as K (F), K (G). F' is on SP6 and contains the key of G and A as K (G), K (A) and G' is on SP7 and contains the key of A and B as K (A), K (B). So the key on a particular cloud is determined by the piece of chunk on the number cloud services provider we placed:

D ((A'+1), (A'+2))

It gives the key of B' and C'. So we can find the key according to the piece of the block we placed on any particular cloud [3]. And the one cloud contains all the encrypted data without keys. Every chunk of data and its key is the division and mirrored affording to RAID 10 employment. A chunk of every data block is divided into two more pieces, and it's replica or copy is also stored on SP1, and the same procedure is also applied on all the cloud services provider sides.

SECURITY

Cloud computing became popular within a couple of years. Security emerges as a most important drawback in cloud computing which affects its popularity, and soon enough it realizes that it is the top most challenge of the cloud computing.

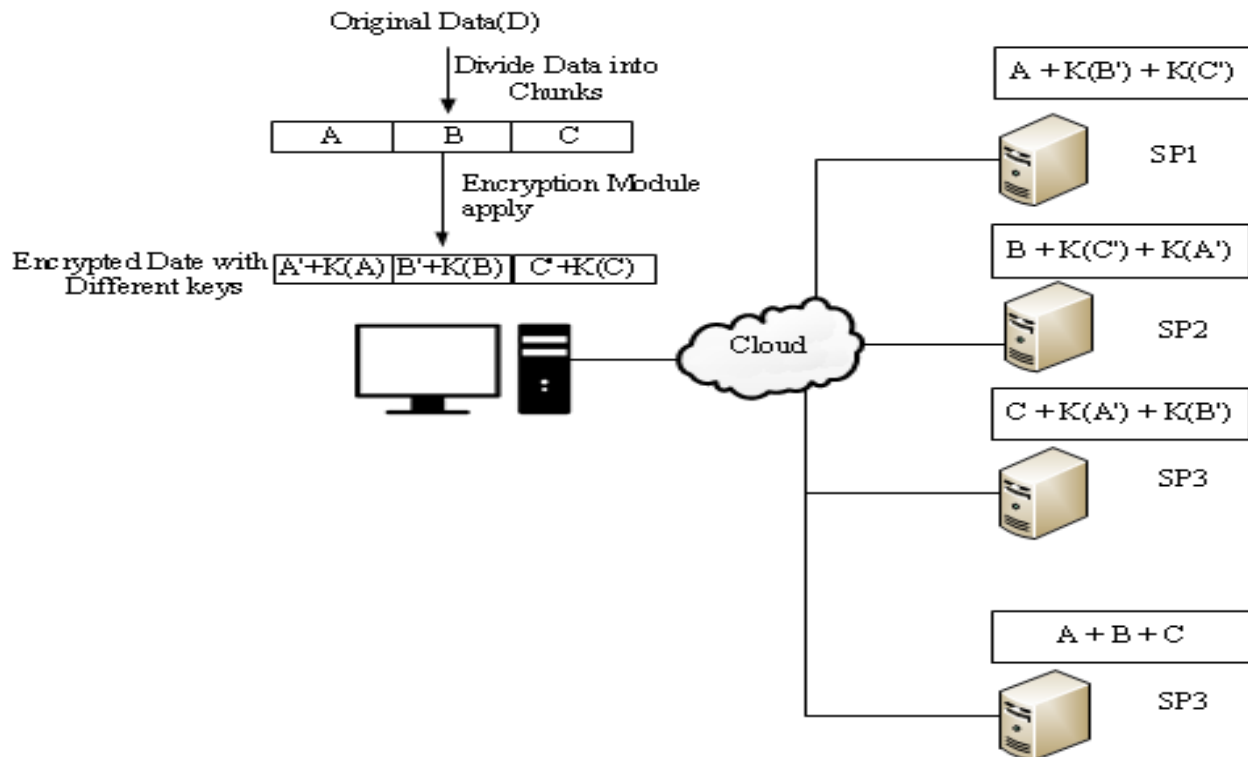


Figure 3: Detail proposed architecture

Besides cloud services provider there is also another source which is also a problem in cloud computing. Moreover, another threat of unauthorized access or intruder also affects the security of the cloud computing [3] [8]. Therefore, to ensure the security we firstly split the data and then offer encryption strategy in our proposed methodology. After this, distribute all the data on different cloud services provider location.

AVAILABILITY

The proposed model promises that the availability of the resources of the data at any time. The services provided by the cloud service provider contain a high level of risk, the risk of single degree of failure, which may destroy all the system. These failures rely upon numbers of factors software, hardware as well as a network failure. Our proposed model gives its solution by splitting and distributing the data on different clouds as compared to store the whole data on the single standalone cloud. If one cloud is down, we can

recover the data from the other cloud which is also available on other clouds [10].

RELIABILITY

Through the key reliability of the offered model can be attained. The chunk of the data cannot be deciphered without a key. Moreover, the intruder did not get the whole data at once. If the particular chunk at the specific cloud is corrupted or lost, it can also be recovered from another cloud as well.

CONCLUSION

The cloud computing became more general, more and more flaws emerge. Currently, the cloud computing is facing several problems, security, reliability and availability are the top most issue of the cloud computing. Security of the sensitive data is the top most priority of an organization. So according to the need, we proposed a better solution to provide more security to the data in the cloud computing. The client's critical data is more secure at the cloud service provider side and the user can access it at any time based on need. Our

model ensures better security, availability as well as reliability.

REFERENCES

- [1]. Y. Singh, F. Kandah and W. Zhang,(2011) "A Secured cost effective multi-cloud storage in cloud computing," IEEE INFOCOM Workshop on Cloud Computing, pp. 619-624.
- [2]. Qi Zhang, Lu Cheng, Raouf Boutaba (2010)" Cloud computing: state-of-the-art and research challenges" J Internet Serv Appl 1: 7–18.
- [3]. Fawaz S. Al-Anzi, Jyoti (2014) et al., "Towards Robust, Scalable and Secure Network Storage in Cloud Computing".
- [4]. Cloud Security Alliance (2011) Security guidance for critical areas of focus in Cloud Computing V3.0.Available: <https://cloudsecurityalliance.org/guidance/csaguide./csaguide.v3.0.pdf>.
- [5]. Zhao G, Liu J, Tang Y, Sun W, Zhang F, Ye X, Tang N (2009) Cloud Computing: A Statistics Aspect of Users. In: First International Conference on Cloud Computing (CloudCom), Beijing, China. Springer Berlin, Heidelberg, pp 347–358.
- [6]. X. Zhang, H. Du, J. Chen, Y. Lin, and L. Zeng,(2011) "Ensure Data Security in Cloud Storage," Proc. International Conference on Network Computing and Information Security (NCIS 11), IEEE Press, pp. 284-287.
- [7]. P. F. Olivera, L. Lima J. Barros and M. Medard, (2010)" Trusted storage over untrusted networks," IEEE Global Telecommunication Conference, pp. 1-5.
- [8]. W. Liu, (2012) "Research on cloud computing security problem and strategy," Proc. 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet 12), IEEE Press, pp. 1216-1219.
- [9]. J. Sun and S.-S Yue, (2011) "The application of Cloud Storage Technology in SMEs," Proc. International Conference on E –Business and E –Government (ICEE 11), IEEE Press, pp. 1-5.
- [10]. P. C. Chen, C. P. Freg, T W Hou and W G Teng,(2011) "Implementing RAID-3 on cloud storage for EMR System," IEEE International Computer Symposium, pp. 850-853.

HSAG-FC: Hybrid of Simulated Annealing with Genetic algorithm and Fuzzy classification for IDS

Mrs. J.Lekha

Research Scholar, Department of Computer Science,
Avinashilingam Institute for Home Science and Higher
Education for Women

&

Assistant Professor, Sri Krishna Arts and Science College
Coimbatore, Tamil Nadu

Dr. G. Padmavathi

Head, Department of Computer Science, Avinashilingam
Institute for Home Science and Higher Education for Women

Coimbatore, Tamil Nadu.

Abstract—Nowadays signature attacks are termed as very big problem because it leads to software vulnerability. Malware writers confuse their malicious code to malicious code detectors such as Signature-based detection. However, it fails to detect new malware. This research article addresses the signature based intrusion detection from Intrusion Detection (IDS) systems. The proposed hybrid techniques for Generation of Signature are done using Genetic Algorithm (GA) and Simulated Annealing (SA) approaches. For this, signature-set in execution statements are selected by using simulated annealing and genetic algorithm, which produce the optimal solution of selection. Then the generated signatures are matched with IDS by using the two pattern matching techniques, namely (i). Finite state automaton based search for Single Pattern matching technique and (ii) Rabin Karp string search algorithm for multiple pattern matching technique. These techniques are used to match the signature as in an effective manner. In addition to this the Fuzzy Logic classification is used to find the degrees of truth of vulnerability for classification. The aim of the proposed work is to improve the final resultant accuracy in compared to existing techniques. The proposed Rabin Karp- fuzzy logic system returns the higher performance metrics namely precision is 88% and Recall is 80% and in open source dataset it contains 30 vulnerabilities this proposed worked well in detecting 28 vulnerabilities/ defect, the accuracy of this proposed is 94.27%.

Keywords: Degrees of truth, Finite state automaton, Fuzzy logic, Genetic algorithms, Intrusion Detection (IDS) systems, Optimization, Signature Generation, Signature matching, Simulated Annealing, Traffic detection.

I. INTRODUCTION

With the help of less computation capabilities and to get sufficient better solution for optimization problem, high-level procedure namely meta-heuristic is used in IDS [1]. There are many evolutionary algorithms that have the heuristic design but especially the genetic algorithm and simulated annealing are more desirable than any other search and optimization algorithm for signature generation and signature matching.

Simulated annealing [2] [3] is a meta-heuristic algorithm which is used to achieve the global optimum of a multi-dimensional function despite the presence of several maxima and minima. It is evolutionary in the sense that unlike traditional optimization techniques such as random walk or hill climbing SA will not get stuck at a local optimum and it is a correspondent method of genetic algorithm for optimization. This method is used to generate the signature in the optimized manner in the gradual processing with time efficiency.

The optimization is done by the random search and also provides the current assignment of values to variables. The Hybrid Simulated Annealing based Genetic Algorithm process is [4] [5] processed by random variable, value selection. This is useful to reduce the number of conflicts. The algorithm accepts the assignment and there is a new current assignment based on the probability depending on the old assignment and how much worse it is than the current assignment.

Genetic algorithm [6] [7] is a high level metaphor of evolutionary biology, is a population method used to solve the optimization problems. Genetic Algorithms generates an entire 'population of them'. Here GA has the five processing view namely population selection, new member creation, Fitness value, crossover and Mutation. Genetic Algorithm is used to read the candidate member populations from the input dataset with the initial population and the retrieved process are ranked from best to worse. This ranked value is used to the corresponding candidate solutions in each population which is created by the iterative population. The three aspects are used to manage the optimization as in successful manner such as rank based population creation, mutation and crossover.

In rank based population creation referred as the new members of the population are created with respect to the highest ranked candidate. The resultant outcome has two alternatives such as the highest ranked candidate solutions or the population to the next generation. These outcomes are

accepted in the crossover. From this outcome, the two parents produce two immediate descendants for the randomly recombined crossover to form immediate descendants. This randomly recombined crossover is avail between the 0.6 and 1.0.

Finally the Mutation usually involves a change in one element in a candidate solution vector from the current generation population to create a solution to the new vector population. The result of the crossover operation is like what would happen if vectors could have crossover. A new child vector whose elements are comprised of some from each of two parents.

With above distinctive character, signature based systems works based on the pattern matching techniques. The IDS contains a database of known-attack signatures like anti-virus and tries to match these signatures with the analyzed data. If the possibilities of attack signatures are found in the IDS, at that time system produce a warning alert. i.e., it returns the false positives against the attacks. This attack effectiveness will go unnoticed until the signature database is updated. So low false positives rate is used to reduce the warning alert of signature database based IDS. But it cannot detect the new attacks without updating the signature database.

To overcome these challenges, the Hybrid Simulated Annealing based Genetic Algorithm techniques are introduced to fulfill those challenges and achieve the new attack detection as well as reduction of false positives and also raise the unclassified alert to the user. This effective hybrid simulated annealing based genetic algorithm processes carry out single pattern matching by finite state automation and multiple pattern matching by Robin Karp string search algorithm. *Moreover the classification is made by the Fuzzy rules [8] on the generated signature. This will optimize and classify the signature database. From the resultant aspect our techniques are effective with respect to the classified patterns on the fuzzy classification systems.* In this research article, the Hybrid Simulated Annealing based Genetic Algorithm techniques which deals with the Genetic Algorithm and Signature matching method[10]; This includes Genetic Algorithm based signature-set Selection, Finite state automation construction [11] and Rabin Karp string search based Pattern matching[12]. In addition to this Fuzzy logic is used for classification.

This paper is organized in section wise. Section II discusses about the literature review. Section III explains the proposed methodology of this work. Section IV describes the observational performance evaluation as well as experimental design. Article conclusion is provided in the final section V with the reference.

II. RELATED WORKS

Gisung Kim et al [13] proposed a technique which hybrid the intrusion detection method that integrates misuse detection model and anomaly detection model in hierarchical structure. The C4.5 decision tree (DT) is used to create the misuse detection model and the support vector machine (1-class SVM) is used to create multiples of anomaly detection models. The proposed method improved the anomaly detection model with known attack information.

David Brumley et al [14] proposed new data-flow analysis for automatically generating vulnerability signatures. The main aid is a new class of vulnerability signature, to exploit successful hijacks control of the program. A vulnerability signature matches a set of inputs which satisfy a vulnerability condition in the program. These techniques are also used in Turing machine signatures, constraint symbolic signatures, and last in regular expression. This proposed techniques measures with known attacks.

David Brumley et al [15] proposed new techniques called weakest preconditions (WP) for automatically generating sound vulnerability signatures which results in fewer false negatives instead of using program binary analysis. The key problem to reducing false negatives, which consider as many as possible different program paths an exploit, may take.

Mabu S [16] proposed a technique called fuzzy class-association rule mining which is based on genetic network programming (GNP) for detecting intrusions in the networks. It optimizes direct graph structures instead of strings in genetic algorithm which leads to enhancing the representation ability with compact programs. By combining fuzzy set theory with GNP, the proposed method is with the varied database such as discrete and continuous attributes for enhancing detection ability. The proposed method is flexibly applied to both misuse and anomaly detection in network problems.

Saddam Khan [17] proposed a classification technique called modified fuzzy with genetic algorithm to analyze students' data. Hybrid of fuzzy and genetic algorithm is employed to optimize the indices entropy and gini index. While running genetic algorithm proper bias values are given for assigning weights to entropy and gini index. Thus optimization classification is evaluated.

The observations due to the literature are tabulated and summarized in the following table 1.

TABLE 1.COMPARISON OF EXISTING DETECTION METHOD

Year	Authors	Techniques Used	Detection mechanisms	Parameters used for evaluation
2014	Kim, Gisung, Seungmin Lee, and Sehun Kim	C4.5 decision tree (DT),support vector machine	Anomaly Detection Model	High accuracy and better performance on known information.
2014	Mabu S, Chen C, Lu N, Shimada K, Hirasawa K	Fuzzy Class-Association Rule Mining, Genetic Network Programming (GNP)	It form Direct graph structures instead of strings in genetic algorithm which leads to enhancing the representation ability with compact programs algorithm.	Better Performance, throughput.
2011	Khan Saddam	Genetic Algorithm Fuzzy rules	Database with a number of records given, also a set of classes is a given, then the problem of classification is to find the class that contains given record	detection accuracy,
2007	David Brumley, Newsome J, Song D, Wang H, Jha S	Weakest preconditions (WP)	Automatically generating sound vulnerability signatures which results in fewer false negatives instead of using program binary analysis.	Reducing false negatives
2006	David Brumley, Newsome J, Song D, Wang H, Jha S	Turing machine signatures, symbolic constraint signatures, and regular expression signatures	A vulnerability signature matches a set of inputs which satisfy a vulnerability condition in the program	High Accuracy, the minimization of vulnerability

From the above table 1, different methods have been proposed to detect the intrusion which affects the applications. From the observations, it is found that many intrusions are found in the applications. To overcome the above limitations, the proposed approach detects and classifies the intrusions in the application.

III. PROPOSED METHODOLOGY

In this paper proposed hybrid simulated annealing and Genetic algorithm (HSAG) methodology for generating signatures and fuzzy logic is used for classification of vulnerabilities. HSAG is an optimization algorithm which is applied to generate the accurate signatures with the evaluated fitness value. Then Finite Automata (FA) and Rabin Karp methodology is used to match the generated signature with online application signatures. This will calculate the accurate

presence of the signature. According to the result vulnerability and non-vulnerability type of applications are found in the network. The flow diagram of the proposed methodology is given below in Fig .1.

A. Hybrid Simulated Annealing based Genetic Algorithm for signature generation (HSAG)

The hybrid approach of Simulated Annealing and Genetic Algorithm is to generate signatures that represent all kind of patterns (attacks in the optimized manner in the gradual processing with time efficiency). The initial process is the simulated annealing and is used to generate the suspicious signature when they climb in the online running application and the next process is the searching that mimics the process of signature generation. HSAG utilizes signature probabilities to help signatures being trapped in local solutions.

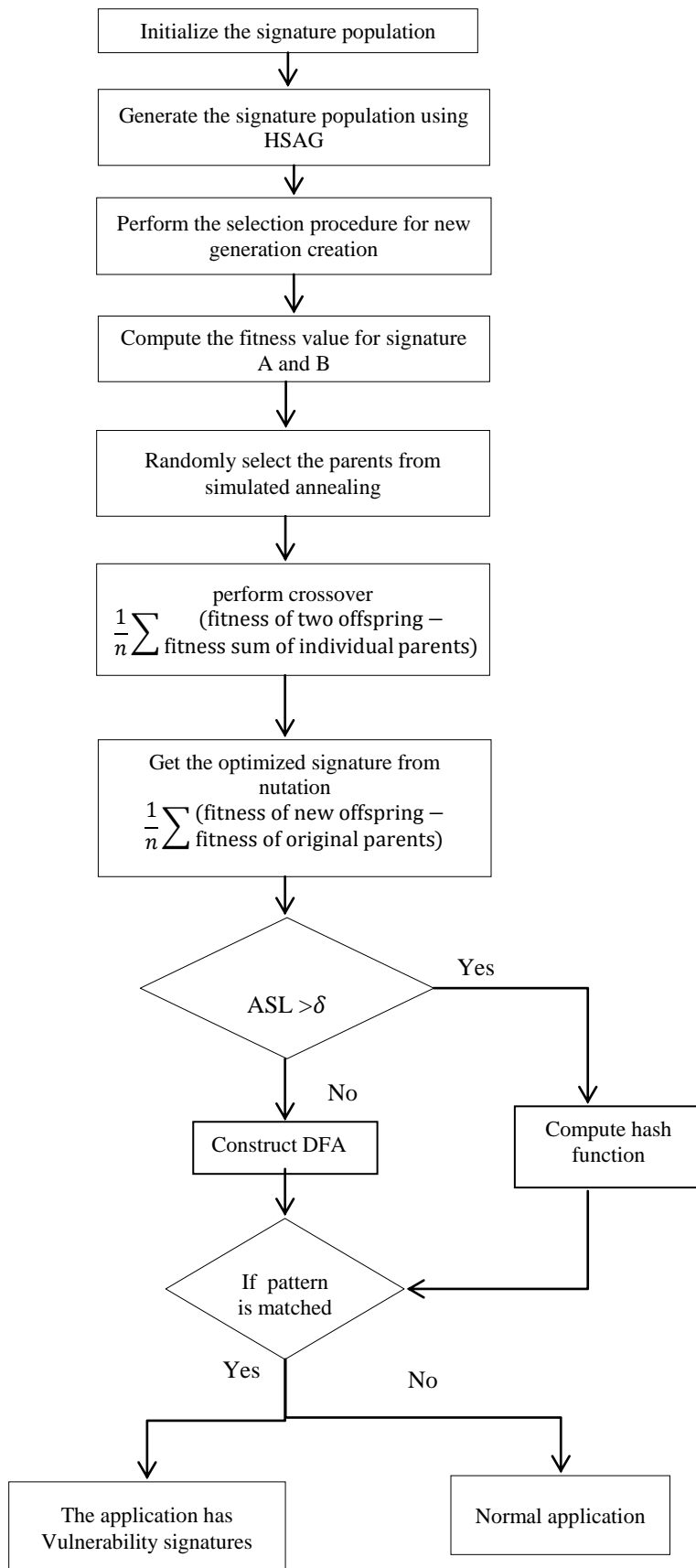


Figure 1Flow diagram of proposed work

Genetic Algorithm (GA) is a heuristic search technique that belongs to the evolutionary algorithm and it imitates the natural selection. To a certain search problems the optimistic solution can be obtained by applying the GA. Based on genetic structure and behavior of chromosomes GA performs the operations in a population.

The GA starts to execute with a set of population. A new population is formed by taking a solution of another population. The newly generated population should be better than the old one. Based on their fitness the solution is taken as an input to form the new solution. The possible solution will be provided by performing the mutation, selection and crossover.

Simulated annealing is a meta-heuristic algorithm which is used to achieve the global optimum of a multi-dimensional function despite the presence of several maxima and minima. It is evolutionary in the sense that unlike traditional optimization method such as random walk or hill climbing it will not get stuck at a local optimum and the simulated annealing is a correspondent method of genetic algorithm for optimization. This method is used to generate the signature in an optimized manner in the gradual processing with time efficiency. The optimization is done by the random search and also provides the current assignment of values to variables. The simulated annealing process is processed by random variable, value selection. This is useful to reduce the number of conflicts, the algorithm accepts the assignment and there is a new current assignment based on the probability depending on the assignment and how much worse it is than the current assignment.

1. HSAG

A combination of simulated annealing with high level metaphor population method such as genetic algorithm is used to solve the optimization problems. Genetic Algorithms generates an entire 'population of them' with the help of population selection, new member creation, Fitness value, crossover and Mutation. These Genetic Algorithms processing has following steps to generate the population and optimization.

Step 1: Reads the candidate member populations from the online application with the initial population and the retrieved process.

Step 2: Iterate population for ranking.

Step 3: The retrieved process are ranked from best to worse.

Step 4: Finds the fitness value in each population with the help of ranked values.

Step 5: Create Rank based population and applies mutation and crossover

The five processing aspects are used to manage the optimization as in successful manner such as rank based population creation, mutation and crossover.

a. Hybrid

A HSAGA for signature that generates number of signature with the highest fitness value $\left[\frac{1}{\epsilon} S_1 + C_1\right] \bar{n}_1$ between the signature that are generated during the crossover and mutation process where S_1 = Signature type and $\frac{1}{\epsilon}$ = step size parameter.

b. Selection

In rank based population creation referred as the new members of the population are created with respect to the highest ranked candidate. The resultant outcome has two alternatives such as the result of highest ranked candidate or it result next generation candidate. These outcomes are accepted in the crossover. From this outcome, the two parents produce two immediate descendants for the randomly recombined crossover to form immediate descendants. The Mutation usually involves a change in one element in a candidate solution vector from the current population that creates new solutions in the next population. The proposed HSAG is used to generate the signature in an optimized manner in the gradual processing with time efficiency.

c. Fitness value

Fitness value is evaluated to find the combinational level in identifying the deed. The fitness value is calculated by relaying the vulnerability signature (A) and normal signature (B) as follows,

$$fitness = \frac{A}{total\ of\ vulnerability} - \frac{B}{number\ of\ normal\ signatures}$$

d. Cross over

Crossover provides a solution by taking more than one parent solution as input and gives a child solution. Parents are selected randomly from the simulated annealing and crossover procedure is performed to create nextnew generation.

$$Crossover = \frac{1}{N} \sum (fitness\ of\ two\ offspring - fitness\ sum\ of\ individual\ parents)$$

N=number of signatures

e. Mutation

The final genetic operator is mutation. It can create a new genetic material in the population by assigning to every individual according to its fitness value.

$$Mutation = \frac{1}{N} \sum (fitness\ of\ new\ offspring - fitness\ of\ original\ parents)$$

```

Begin
2:   Read all signature
3:   Initialize the parameters of signature
4:   Generate signature population
5:   based on the Initialized
parameters
6:   Set the limits of generated signature
7:   From initial stage to final
end of
8:   signature sequences
9: begin Case 1:
10:  if Total number of read signature < generated
signature
11:    Generate new signature
12:    Execute till end of signature
13:    Collect out of resultant
signature
14:    do apply: simulated annealing
15:    Reads simulated annealing outcome
16:    Apply: genetic algorithm
17:    Compute the fitness function
18:    To each individual outcome
of
19:    Simulated
annealing
20:    while Condition: If fitness function is high
21:    for any simulated annealing
22:  go to step 10
23: end case 1
End

```

Algorithm 1: Hybrid Simulated Annealing based Genetic for signature generation (HSAG)

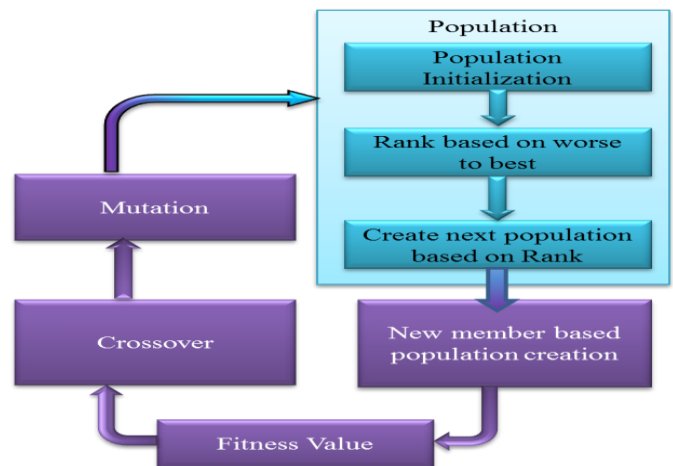


Figure 2. Hybird Simulated Anealing with Genetic Algorithm

2. Signature Matching Techniques

The hybrid simulated annealing based genetic algorithm process carries two types of pattern matching for effective solution. Selection of techniques is based on average signature length. If the average signature length is below threshold value “single pattern matching” technique is carried out by using Finite State Automata techniques or else average signature length is greater than threshold value “multiple patterns matching” technique is carried out by Rabin Karp String Search Algorithm.

a. Single pattern matching

For single pattern matching, Deterministic Finite Automata (DFA) is used and it is defined as machine that accepts or rejects the finite language which produces a unique computation of the automata on every input string and it is applied directly. In pattern matching scheme, input signatures are matched with the online applications by using regular expression patterns. Exhaustive and Non-Overlapping matching styles are executed in the online applications to find start and final positions.

In Non-Overlapping approach, for the matching process, let S be a function from a pattern A and a string W to a power set of P' . The matching process will result all non-overlapping substrings that matched the pattern appearing in multiple locations from the input string. Non-Overlapping matching for matching the signature provides better analysis of intrusions found in the online application. This matching lacks in memory-efficient DFA. One Pass Search execution process is executed explicitly by DFA to handle pattern substring matching. In this proposal, DFA uses non-overlapping matches and one pass search. Figure 5.5 below illustrates DFA for regular expressions $ua?xb*yc.zd$

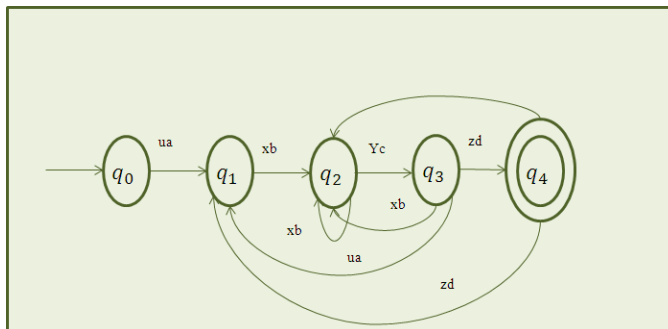


Figure 3. Construction of DFA

The detection step above clearly discussed the intrusion detection using DFA. After the detection of signature, they are classified as vulnerabilities using Fuzzy Logic classifier. The algorithm for constructing DFA is given below

```

Input: signature S
Output: signature matched
Begin
  For each S transfer
    Scan and compare S with regular expression
    pattern  $P_S$ 
    If (S matches  $P_S$ ) Then
      Divide S into sub packets s
    End if
    For m=1 to n
      Construct ps (m) into state qmi) using transition
      function  $\delta$  Compress stateless
      rs(m)=encode(q(m))
      q(m)=decode(rs(m))
      Compute decode latency
    End for
    For m=1 to n
      For k=1 to n
        Compare(s(m),s(k))
        If(s(m)matches(k))Then
          Count=count+1
        End if
      Until end of k
    Until end of m
    If (count>threshold) Then
      Signature matched
    End if
  End for
End
  
```

Algorithm 2: Deterministic Finite Automata (DFA) for single pattern matching

b. Multiple Patterns Matching

Multiple pattern matching is done using Rabin-Karp approach is a string searching algorithm that uses hashing to find any one of a set of pattern strings in a text. During preprocessing, for all patterns hash values are calculated and these values are stored in an ordered table. Matching of signatures are done by calculating the hash value for each signature in the execution statements and the ordered table it search for hash value by using binary search. If a hash value is found to be matched, the corresponding signature is compared with the signature in found in the applications. Below fig 4 shows an example for multiple pattern matching technique.

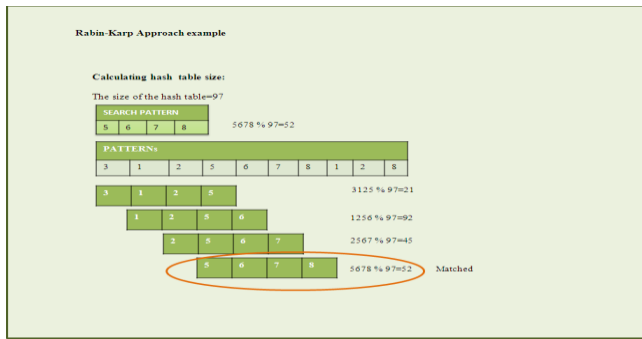


Figure 4. Example of multiple string matching using Rabin-Karp Method

After the detection of signature, they are classified as vulnerabilities using Fuzzy Logic classifier. Algorithm for Rabin-Karp method is described below.

Algorithm:

Input: signature detected pattern , generated signatures

Output: Signature detected

Begin

1 detected pattern dp[1..n], generated signatures sign [1..m]

2 hpattern := hash(dp[1..m]); hs := hash(sign[1..m])

3 **for** i from 1 to n-m+1

4 **if** hs = hpattern

5 **if** sign[i..i+m-1] = dp[1..m]

6 **return** i

7 **end if**

8 **end if**

9 hs := hash(sign[i+1..i+m])

10 **end for**

11 **return** not found

End

Algorithm 3: Rabin-Karp approach for multiple pattern matching

c. Fuzzy classification

The intrusion detection is a two-way classification problem; the objective is to classify the signatures in the execution statements in two categories (vulnerability and normal), using signatures the matched patterns are classified as vulnerability and rest is known as resistance; the classifications are done with fuzzy rules based on fuzzy logic concepts.

Fuzzy rules have the form:

IF state **THEN** subsequent [load]eqn(1)

Where, State-is a complex fuzzy expression

Subsequent- is an expression as atomic, and

Load -is a real number which assures the confidence of the rule.

The identified signatures are determined as j+1 class in

which each and every object (signatures) is classified using fuzzy classification. These classes are divided into vulnerability class (intrusions or attacks) and normal class. The data set consists of a set of signatures with n+1 attributes. The object characteristics are determined by first n attributes and last attributes is used to determine the class that the objects belong to. The maximum technique is used to classify the signature as the class in the resulting portion of the rule that has the maximum truth-value (TT).

class

$$= \{R \text{ if } TT(F_R) > TT(F_{V-k}) \forall k$$

$$= 1 \dots i \dots V_k \text{ if } TT(F_R) < TT(F_{V-k}) \\ TT(F_{V-m}) < TT(F_{V-k}) \forall m = 1 \dots k \ m \neq k$$

.....eqn(2)

Where,

R represents the normal class

V_k Represents the Abnormal class

F_R is the rule for the normal class

F_V is the rule for the k^{th} abnormal class

Conditions for fuzzy classification as follows,

Normal: **IF** x is **HIGH** and y is **LOW THEN** pattern is normal [0.4]

Abnormal: **IF** x is **MEDIUM** and y is **HIGH THEN** pattern is abnormal [0.6]

$Abnormal_k$: **IF** x is **LOW THEN** pattern is $Abnormal_k$ [0.6].

According to the set of fuzzy values, the vulnerability applications and normal applications are classified.

Below algorithm 4 describe the pseudo code for proposed methodology. Determination of malicious applicant in an efficient manner is the objective of this work. The working principle is clearly showed in the flow chart to obtain the vulnerability. The clustering is performed by the modified Fuzzy logic algorithm. From the cluster the appropriate vulnerability is classified.

Pseudo code:

Input: Online Applications

Output: Classification of Vulnerability applications and non-vulnerability applications

Begin

Read all signature

Initialize the parameters of signature

Generate signature population based on the Initialized parameters

Set the limits of generated signature

From initial stage to final end of

 Signature sequences

/hybrid simulation annealing and genetic algorithm****

Case 1:

if Total number of read signature < generated signature

Generate new signature

 Execute till end of signature

 Collect out of resultant signature

do apply: simulated annealing

 Reads simulated annealing outcome

 Apply: genetic algorithm

Compute the fitness function

 To each individual outcome of

 Simulated annealing

While (If fitness function is high

for any simulated annealing)

 go to step 10

End while

End if

End case 1

Case 2: /Finite Automata****

For each S transfer

 Scan and compares with signature generation

If (S matches S_a) **Then**

 rs(m)=encode(q(m))

 q(m)=decode (rs(m))

 Compute decode latency

Form=1 to n

Fork=1 to n

Compare(s(m),s(k))

If(s(m)matches(k))**Then**

 Count=count+1**endif**

Until end of k

Until end of m

) Then

If(count>threshold)**Then**

 Signature matched

End if

End for

End case 2

Case 3: /Rabin Karp method****

 Pattern(Detected pattern dp[1..n], generated

 signatures sign [1..m]))

 hpattern := hash(dp[1..m]); hs :=

 hash(sign[1..m])

for i from 1 to n-m+1

if hs = hpattern

if sign[i..i+m-1] = dp[1..m]

return i

end if

end if

 hs := hash(sign[i+1..i+m])

end for

End case 3:

Case 4: / fuzzy classification****

for k=1 to i

for m=1 to k

if $TT(F_R) > TT(F_{V-k})$

Return R

End if

if $TT(F_R) < TT(F_{V-k}) \wedge TT(F_{V-m})$
 $< TT(F_{V-k})$

Return V_k

End if

End for

End for

End case 3

End

Algorithm 4: proposed HSAG-FC: Hybrid of Simulated Annealing with Genetic algorithm and Fuzzy classification for IDS

IV. EXPERIMENTATION RESULTS

This section gives brief description about the dataset used for this technique and the statistical analysis of the results obtained by this technique which is compared with some existing works. The proposed methodology is implemented using Java and SQL server. This work takes three open source applications for vulnerability detection, namely, Tomcat 3.0, Tomcat 3.2.1, and Jigsaw 2.0. These dataset are collect from www.cvedetails.com/vulnerability-list/vendor_id-45/product_id-887/version_id-2591/Apache-Tomcat-3.0.html, https://www.cvedetails.com/vulnerability-list/vendor_id-45/product_id-887/version_id-5532/Apache-Tomcat-3.2.1.html, and https://www.cvedetails.com/vulnerability-list/vendor_id-45/product_id-887/version_id-5532/Apache-Tomcat-2.0.html.

[3.2.1.html](http://www.cvedetails.com/version/18232/W3C-Jigsaw-2.0.html), <http://www.cvedetails.com/version/18232/W3C-Jigsaw-2.0.html>. The proposed work is done in three main phases of signature generation, signature matching and classification. The performances metric are evaluated with the existing system for classification of vulnerabilities applications in the online. The performance results such as Precision, Recall, Accuracy, are listed in below table 1.

This paper deploys set of signatures to measure the number of false alarms. In the false alarms, the false negative is set when undetected exploits are happened and false positive results in incorrect labeling from safe execution. The test cases are flagged as attacks by monitoring the output of IDS. The collection of vulnerabilities is actually checked with the set of signatures that are already generated.

A. Evaluation metrics

The performance of this work is measured using Precision, Recall, Accuracy which shows that an efficient result towards the proposed protocol. Those results are discussed briefly below.

Table 2QoS parameter for proposed HSAG-FC

Parameters	Existing system	Proposed System	Improvement
Precision	85	88	3.5294%
Recall	78	80	2.5641%
Accuracy	86.67	94.278	8.7781%

Table 1 shows the classification accuracy of the proposed and existing technique. The proposed technique achieves higher accuracy for both dataset while compared with existing technique.

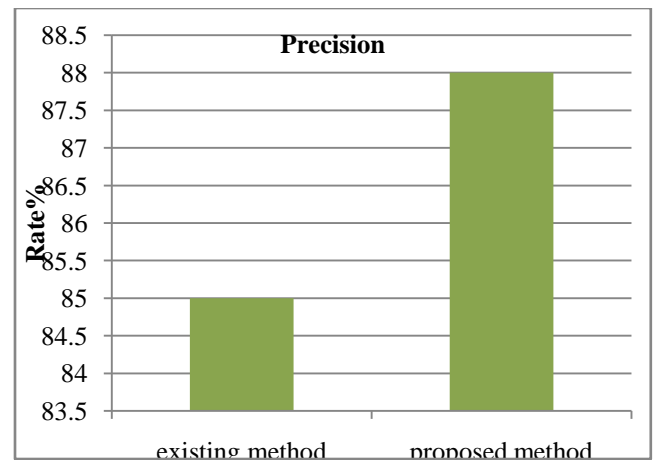


Figure 5. Precision of proposed methodology

Figure 5 illustrates the performance measure of precision for existing Genetic algorithm and proposed hybrid of simulated annealing and genetic algorithm. In the open source application, hybrid of simulated annealing and genetic algorithm are 3.5294% higher than the existing system.

Precision is defined as the ratio of the number of regained vulnerability in classification to the number of retrieved dataset..

Precision

$$= \frac{\text{number of regained relevant in classification}}{\text{number of retrieved}}$$

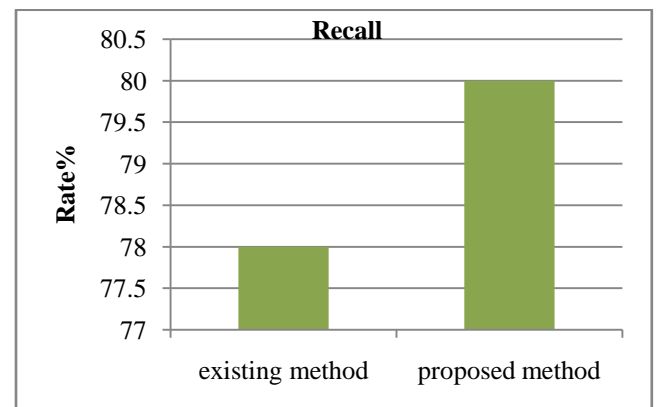


Figure 6 .Recall of proposed methodology

Figure 6 illustrates the performance measure of recall for existing Genetic algorithm and proposed hybrid of simulated annealing and genetic algorithm. In the open source application, hybrid of simulated annealing and genetic algorithm are 2.5641% higher than the existing system.

Recall is defined as the ratio of the number of regained vulnerabilities in classification to the total number of relevant in the dataset.

Recall

$$= \frac{\text{number of regained vulnerability in clasifcation}}{\text{total number of relevant in the dataset}}$$

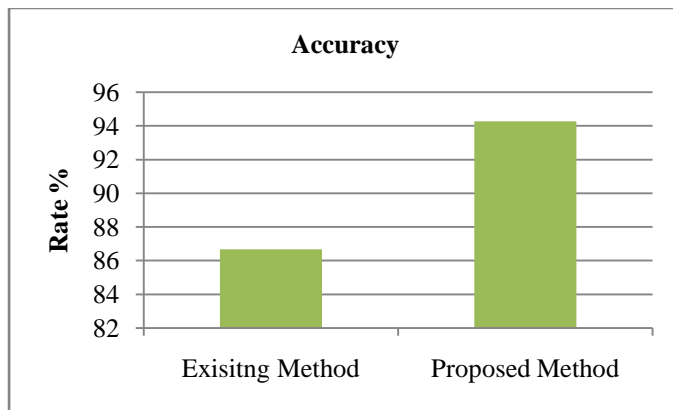


Figure 7. Accuracy of Proposed Methodology

The above Figure 7 illustrates the performance measure of accuracy for existing Genetic algorithm and proposed hybrid of simulated annealing and genetic algorithm. In the open source application, hybrid of simulated annealing and genetic algorithm are 5.7616% higher than the existing system.

With reference to the fuzzy logic cluster properties, multiple parts of the input data that shares common label of same cluster percentage called accuracy of the clusters. This accuracy is calculated by the small clustering inputs and the whole cluster are calculated based on the results of small clustering inputs.

V. CONCLUSION

This work proposes an approach to application-based intrusion detection relying on hybrid of simulated annealing and genetic algorithm. Using this type of signatures, it will define the conditions of attacks, format mismatch or discontent by exploit, by only referring to the set of execution elements that occur during an attack. The proposed solution is applicable of online failure detection, which aims to prevent a system from an attack. This paper proposed hybrid optimization and classification based intrusion detection is to detect the vulnerability in the network. Fuzzy classification based hybrid of simulated annealing and genetic algorithm is used to classify the unknown attack. Thus, the proposed method has been implemented as real-time application. The experimental results of proposed method is shown in section IV that gives better detection accuracy of 94.278% in application vulnerabilities when compared to that of existing approach.

REFERENCES

- [1] Dharmapurikar, Sarang, and John W. Lockwood. "Fast and scalable pattern matching for network intrusion detection systems." *Selected Areas in Communications*, IEEE Journal on 24.10 (2006): 1781-1792.
- [2] Bryan, Kenneth, Pádraig Cunningham, and Nadia Bolshakova. "Application of simulated annealing to the biclustering of gene expression data." *Information Technology in Biomedicine*, IEEE Transactions on 10.3 (2006): 519-525.
- [3] Mantawy, A. H., Youssef L. Abdel-Magid, and Shokri Z. Selim. "Integrating genetic algorithms, tabu search, and simulated annealing for the unit commitment problem." *Power Systems*, IEEE Transactions on 14.3 (1999): 829-836.
- [4] Chen, Yaow-Ming, Chien-Hsing Lee, and Hsu-Chin Wu. "Calculation of the optimum installation angle for fixed solar-cell panels based on the genetic algorithm and the simulated-annealing method." *Energy Conversion*, IEEE Transactions on 20.2 (2005): 467-473.
- [5] Deneshmand, Shabnam, and Hossein Heydari. "A Diversified Multiobjective Simulated Annealing and Genetic Algorithm for Optimizing 3-Phase HTS Transformer."
- [6] Yuen, Shiu Yin, and Chi Kin Chow. "A genetic algorithm that adaptively mutates and never revisits." *Evolutionary Computation*, IEEE Transactions on 13.2 (2009): 454-472.
- [7] Huang, Shih-Chia, Ming-Kai Jiau, and Chih-Hsiang Lin. "Optimization of the Carpool Service Problem via a Fuzzy-Controlled Genetic Algorithm." *Fuzzy Systems*, IEEE Transactions on 23.5 (2015): 1698-1712.
- [8] Chang, Yu-Chuan, Shyi-Ming Chen, and Churn-Jung Liao. "Fuzzy interpolative reasoning for sparse fuzzy-rule-based systems based on the areas of fuzzy sets." *Fuzzy Systems*, IEEE Transactions on 16.5 (2008): 1285-1301.
- [9] Pancho, David P., J. Marcos Alonso, Oscar Cordon, Arnaud Quirin, and Luis Magdalena. "FINGRAMS: visual representations of fuzzy rule-based inference for expert analysis of comprehensibility." *Fuzzy Systems*, IEEE Transactions on 21, no. 6 (2013): 1133-1149.
- [10] Zhu G, Zheng Y, Doermann D, Jaeger S. Signature detection and matching for document image retrieval. *Pattern Analysis and Machine Intelligence*, IEEE Transactions on. 2009 Nov;31(11):2015-31.
- [11] Ottavi, Marco, Gian-Carlo Cardarilli, D. Cellitti, Salvatore Pontarelli, Marco Re, and Adelio Salsano. "Design of a totally self checking signature analysis checker for finite state machines." In *Defect and Fault Tolerance in VLSI Systems*, 2001. Proceedings. 2001 IEEE International Symposium on, pp. 403-411.
- [12] Dayaratne N, Ragel R. Accelerating Rabin Karp on a Graphics Processing Unit (GPU) using Compute Unified Device Architecture (CUDA). In *Information and Automation for Sustainability (ICIAFS)*, 2014 7th International Conference on 2014 Dec 22 (pp. 1-6).
- [13] Kim, Gisung, Seungmin Lee, and Sehun Kim. "A novel hybrid intrusion detection method integrating anomaly detection with misuse detection." *Expert Systems with Applications* 41.4 (2014): 1690-1700.
- [14] Brumley D, Newsome J, Song D, Wang H, Jha S. Towards automatic generation of vulnerability-based signatures. In *Security and Privacy*, 2006 IEEE Symposium on 2006 May 21 pp. 15.
- [15] Brumley D, Wang H, Jha S, Song D. Creating vulnerability signatures using weakest preconditions. In *Computer Security Foundations Symposium*, 2007. CSF'07. 20th IEEE 2007 Jul 6 pp. 311-325.
- [16] Mabu S, Chen C, Lu N, Shimada K, Hirasawa K. An intrusion-detection model based on fuzzy class-association-rule mining using genetic network programming. *Systems, Man, and Cybernetics, Part C: Applications and Reviews*, IEEE Transactions on. 2011 Jan;41(1):130-9.
- [17] Khan, Saddam. "Analyzing students' data using a classification technique based on genetic algorithm and fuzzy logic. IEEE, " *Computing, Communication & Automation (ICCCA)*, 2015 International Conference on. 2015.

A Novel Technique for Jobs Scheduling In Cloud Computing Systems

Muneer Bani Yassein*, Yaser Khamayseh and Ali Hatamleh

Department of Computer Science, Jordan University of Science and Technology
Irbid, Jordan,

Abstract—Recently, cloud computing has occupied a large place in the world, especially in the field of information technology. It is characterized as mainly rely on the Internet to provide services for organizations and consumers and to take advantage of resource sharing, in addition to that it is associated with many of the central remote servers to maintain user data, so it has become an effective way that will allow the world to use the many kind of applications without making an effort to be downloaded. Many job scheduling algorithms have been proposed to achieve both customer satisfaction and high resource utilization. However, better algorithms to achieve these goals efficiently are still needed. This paper proposes a hybrid technique for jobs scheduling based on Neural Network (NN) algorithm. The proposed algorithm classifies the jobs into four different classes. Furthermore, a Heuristic Resource Borrowing Scheme (HRBS) is proposed to exploit all services which has offered by cloud computing. Simulation is conducted using extensive (Cloud-Sim) simulator to measure the efficiency of the suggested algorithm in terms of average throughput, average turnaround time and average of context switch. The obtained results show that the proposed scheme outperforms other state of the art scheduling schemes.

Keywords—Cloud Computing, Job Scheduling, Hybrid Technique, Virtualization.

I. INTRODUCTION

A. Cloud Computing Overview

Cloud computing is a parallel distributed system which includes many servers distributed in a different geographic area these servers are connected with Internet. In addition to that various tasks require to be implemented by using the obtainable resources to perform high performance, minimal response time, and fully utilization of resources [1].

Cloud computing has ability of providing many applications with the required services dynamically by scaling up/down. Meaning, cloud computing uses resources sharing to provide resources for customers to process their jobs based on the agreement between customers and server providers. Scaling service can be divided into two types: the first one is called predictable (by access pattern through night/day), the second one is called unpredictable (by little increase in the application services) [2]. These features can be more important for flexible information such that web hosting.

Cloud computing has ability for working in parallel system, so it can be containing huge number of applications. It has many aspects including job scheduling [2]. Now cloud computing has been introduced in the establishment of most of the areas of computing technology for instance, storage system, networking, Service-Oriented Architecture (SOA), Service-Level Agreement (SLA), Quality of Service (QOS), and Business Process Management (BPM) [2]. However, cloud computing began to face a challenge in terms of their understanding deeply.

B. Cloud Computing Service Models

Cloud computing has been classified into three categories based on service models see figure 1 [6].

- Infrastructure as a Service (IaaS): This model provides the physical equipment which the user requires, such as servers, storage, and virtual machines networks.
- Platform as a service (PaaS): This model provides a computing platform such as, programming language executing environment, operating systems and database.

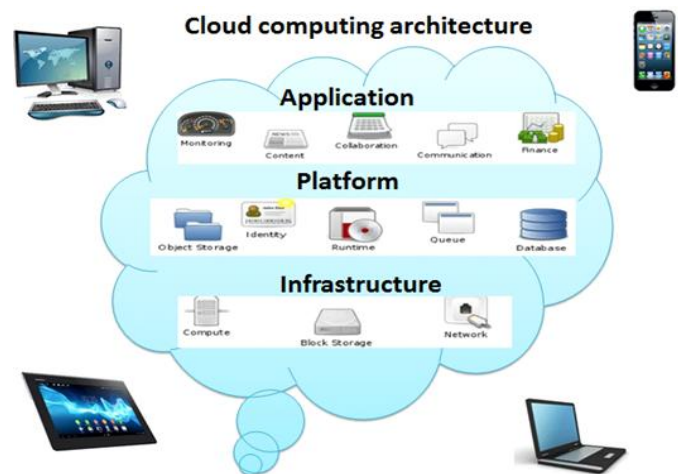


Figure 1. Cloud computing architecture [6].

- Software as a service (SaaS): This model allows consumers to use already installed software

applications such as environment, programming language execution, database, and web server, without the need to buy or manage the software applications.

C. Deployment Models (cloud infrastructure types)

In this section we present the common cloud infrastructure types [6]:

- Public cloud: in this type, the service providers of cloud computing make cloud computing resources, storage and cloud applications available to the public users for free as strategy of pay-as-you-go. Usually, public cloud service providers such as: Microsoft, Amazon AWS and Google own and operate infrastructure and only provide access via the Internet [7].
- Private cloud: in this type, cloud is operated by a single organization, and it is often managed internally or externally by a third-party. Furthermore, it can be hosted internally or externally [14]. There are several examples of the private cloud infrastructure such as Nimbus [13], Open Nebula [12], or Eucalyptus [12].

Hybrid cloud: in this type of cloud emerged from the combination of two or more of the public and private cloud, which offers the benefits of different infrastructure types [5]. Community cloud: this infrastructure characterized as shared between various corporations with a particular community based on some joint aspect like security.

- This type of infrastructure is shared by several organizations with a specific community depends on common concerns such that security. Community cloud costs are split by these organizations, while the private cloud costs are carried only by one organization. Furthermore, it can be managed by those organizations or by a third party [8]. these companies have authorized to administrated it.

D. Virtualization

Virtualization: is a technique which has ability to run more than one virtual machine at the same time related one machine [8]. In fact, when a customer wants to use cloud computing, a virtual machine is rented by the customer that is responsible for gathering the needed resources to execute the customer's job. In addition to that, we can list many advantages for resource sharing:

- High utilization for machine's resources due to using many isolated operating systems.
- Resources sharing enables the use of different operating systems simultaneously (e.g. Windows and Linux) [6].

So we can have defined hypervisor as a software layer which has ability to connect various virtual machines with underline physical resource. In addition to that operating systems have authorized to make access to these resources. However, in virtualization the operating systems have ability to access to hardware by using this layer. The software layer (hypervisor) carried out many instructions rather than aids of virtual machines, the main common technology which

implement virtualization technique are XEN, KVM, and VMWARE [20]. The virtualization technique has many advantages in many areas especially in cloud computing regarding to: Reallocating virtual machines which enable to minimizing number of resource usage & number of physical machine due to converting virtual machine to tiny physical machine. In other words, this is called server consolidation [6].

Migration of virtual machines is a crucial issue that is supported by virtualization. Migration process transfers the main memory pages and states of the current virtual machine to the target machine. Migration policy dictates how this migration of virtual machines takes place between physical machines [21]. For example, when there is a high load on the physical machine or, when there aren't adequate hardware resources for a certain virtual machine, the virtual machine is migrated to other physical machine with sufficient hardware resources. Maintaining performance level, the placement policy, and the migration policy of virtual machine are crucial challenges of virtualization. For solving many problems, we have developed new approaches and policies.

The residue of the paper has been divided as follows: Section II introduces the Literature Review. Section III discusses the proposed job Scheduling technique. Section IV presents the simulation results that were generated using Cloud-Sim Simulator. Finally, Section V explains the conclusion of paper.

E. Jobs Scheduling in Cloud Computing System

Job scheduler main job is to distribute costumer's jobs across the required resources to process them. There are different types of job schedulers which are based on different criteria, some of them are static, dynamic, and centralized job schedulers. Many distributed scheduling standards can be classified as mentioned in [15], [16] and [17]: Static Scheduling: this technique is called pre-scheduler jobs, the main idea of this scheduler is gathering information with respect to resources and jobs which exist in application must be well known in advance. Furthermore, a job has enabled to assign for available resource, and it stays assigned to this resource until it is finished processing, consequently its suitable if rely on scheduler's perspective [15].

Dynamic Scheduling: is a technology which the jobs are dynamically obtainable with time by using specific scheduler, thus it has become known about the run time for these jobs. In addition to that, it is more efficient than static but it is hard to include load balancing factor for taking efficient & static scheduling algorithm [15].

Centralized Scheduling: is responsible for taking decisions about which jobs are supposed to be assigned to which resources. By using centralized scheduling, we gain more efficiency& control, ease of implementation and monitoring the resources. However, the scheduler suffers from several measures such as: scalability, fault tolerance, efficient performance, and single point failure. So it's not implicit any recommendation for wide grids [15].

Decentralized Scheduling: it is emphasizing on employs several units which make the decisions about which jobs are supposed to be assigned to which resources. This makes it

^{*}Corresponding author: Muneer Bani Yassein)

more realistic for real grids in case of no central control the local schedulers' requests responsible for managing the jobs queue [16]. Also, this type of scheduling does not suffer from a single point of failure.

Co-operative Scheduling: in this type of scheduling, the system has already different types of schedulers, in scheduling process each scheduler has responsible to accomplish some activity rely on known rules and current system user [16] for common systems.

Preemptive Scheduling: it is emphasizing on giving each job in the system permission to stop working through the running step, and then it can be moved to other available resources leaving the previous resource ready for next job. if will be taken into account the priority, more benefit can be obtained from it[16]. For example, if a job that has a higher priority arrives, it may interrupt the job which is using the resource. Also, if there is a greedy process which has been using the resource for a long time, such job can be interrupted so other jobs can use the resource.

Non-Preemptive Scheduling: in which resources aren't being allowed to be interrupted or re-allocated until the running and scheduled job finishes its execution [16]. So in this kind of schedulers, a process cannot be interrupted even if it has been using the resource for a long time.

II. LITERATURE REVIEW

Several algorithms have been proposed to address the problem of job scheduling in a distributed environment (e.g., cloud system); in this section we discuss some of these algorithms:

A. First Come First Serve

This algorithm is responsible for organizing and allocation resources to jobs. It illustrates a way of queue processing. Meaning, the order which comes first gets processed or serviced first. The order that comes second waits in the queue until the first job gets serviced, and then it gets its turn to be served [18].

Schwiegelshohn and Yahyapour explain (FCFS) by using parallel processing to administrating the time of resource allocation with respect to candidate task from all incoming tasks [19]. This is known as Opportunistic Load Balancing (OLB) or myopic algorithm [19]. The main idea of OLB is taking each task available in a queue and assigns it to a specific resource randomly, regardless to execution time using these resources that are mentioned in [12], [11], [5] and [6]. It is obvious that (OLB) works to remain each the resources not available (busy) at all time.

Round-robin (RR) is a scheduling algorithm which implements circular queue for tasks without take into account the priority thus it is easier than other and very simple.

B. Round Robin

In RR, all jobs get a certain portion of time to user resources which is called time slice. Thus the fairness is applied for each job, as Ruay-Shinung Chang supposed in [13].

However, this algorithm is bad choice for jobs that have completely different in their size and requirements.

C. Weighted Round Robin

The Weighted Round Robin (WRR) CPU Scheduling algorithm has the advantages of both the round robin and priority-based algorithms. In other words, Weight-based processes finish faster than the other jobs which boost efficiency for higher priority jobs [3].

D. Randomized Round Robin

Bani Yassein, et al. Proposed the Randomize Round Robin, this algorithm is an enhancement to the traditional RR, to overcome its drawbacks, such as, the high degree of context switching between jobs, low throughput and starvation, where some a greedy process can seize the bandwidth, making the other processes waiting until it finishes the whole job, in which the job size might be too large and needs too much time to be executed while the others are waiting [3].

E. Minimum Completion Time

R. F. Freund et al. proposed Minimum Completion Time (MCT) [19], that focuses on selecting any task randomly then assign it to a suitable resource for execute it and with (MCT)simultaneously at [19]. Consequently, it can be consuming many resources in poorly manner in case some assigned happened without taking (MCT). Based on these criteria of scheduling causes some tasks to be assigned for resources that haven't the Minimum execution time in turn poorly consume some resources.

F. Minimum Execution Time

R. F. Freund et al., produce an efficient algorithm for jobs scheduling which called Minimum Execution Time (MET) algorithm. The basic idea of this algorithm is to choose job to be run on a resource that has minimum execution time without considering the resource availability.

G. Min-Min

Min-Min scheduling is fully dependent on Minimum Completion Time (MCT) which is focuses on selecting the tasks randomly then assign to suitable resource with take into account it has (MCT).

Firstly, a scheduler takes a set of tasks and a set of available resources. Then it assigns tasks with minimum MCT to the next available resources. Such process is repeated after removing job from the map step to become the all (un-mapped) set is empty [10].

H. Max-Min

Similar to Min-min, a scheduler organized tasks by taking the running time of the tasks. Instead of selecting the minimum MCT, the maximum MCT is selected. It focuses on assigning priority to large tasks over others small [15]. The Max-min algorithm does better than Min-min algorithm when the number of short tasks is much more than the long ones.

III. A HYBRID TECHNIQUE FOR JOBS SCHEDULING IN CLOUD COMPUTING SYSTEM

A. Overview

Cloud computing is a parallel distributed system that consists of virtualized computers and large scale of interconnected based on Service-Level Agreements (SLA) which is established through negotiation between the consumers and the service provider [2].

Furthermore, cloud computing, is concerned with the ability of dynamically providing the amount of required resources to satisfy the customers' demands.

Moreover, when a job arrives from a customer, the job at first is distributed by job schedulers across queues that belong to the required resources which are to process such jobs according to the customer's SLA.

In this paper, we propose a scheduling algorithm that receives jobs from the customers and utilizes a Neural Network (NN) component to schedule the incoming jobs. Before scheduling, jobs are normalized according to the SLA parameters. By applying normalization we gain normalization weight value, namely F value. After those jobs are assigned their F values, the job scheduler assigns the jobs to the appropriate classes according to their weights.

Furthermore, we propose a heuristic resources borrowing scheme (HRBS), in which resources are moved from non-active classes to the active ones.

B. Neural Network Algorithm

Neural Network Algorithm is an artificial intelligence algorithm that is used to observe and to learn representations of the input of a training set which capture characteristics of the input distribution.

As learning in neural networks is specifically useful in applications where the complexity of the data or task makes difference of how they are supposed to be handled. Meaning, after the NN learns from the training set, it starts classifying jobs to several classes which we will discuss later in this paper.

C. Scheduler

The job scheduler main task is to distribute jobs which are ordered by consumers across the available resources to process these jobs.

There are different types of job schedulers as mentioned in section 1. We decided to use dynamic, preemptive and decentralized scheduling to fully utilize the system resources and to prevent greedy customers from taking resources for a long time.

D. The Proposed Algorithm

The proposed algorithm is based on Neural Network Algorithm (NN) for distributing jobs to the four weighted classes, based on their computed weights, see Figure 2.

Each of the four classes has a certain priority to execute their jobs, the weight of each class (W_i) is assigned as follows; class A=0.4, class B=0.3, class C=0.2 and class D=0.1. We tried different class weights values in the simulation and found

out that these values achieve a near optimal resource utilization.

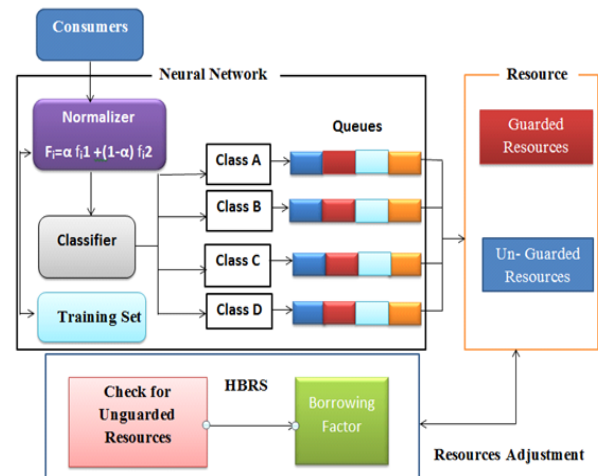


Figure 2. A hybrid technique for jobs scheduling

E. Simulation Environment

All experiments were conducted and developed using Cloud-Sim V 3.0. The Cloud-Sim simulator is implemented using Java JDK 1.7, and it provides support for setting up and simulating a cloud based servers (data centers) environments, including dedicated management interfaces for virtual storage and processing capabilities. In the simulation scenarios we used the parameters that are shown in Table I.

TABLE I. SIMULATION PARAMETERS

Parameter	Value
Algorithm	RR, WRR, RRR
Number of VM	5
Number of data centers	3
Queue strategy	Dynamic memory allocation
VM RAM	512 MB
File Size	300 MB
Host Ram	16 GB
Operating system for data center	Linux
Bandwidth	1000 Mbps
Number of CPU	1

The experiments were carried out to evaluate the performance of the proposed algorithm against the traditional RR, WRR, and RRR in terms of the following:

- Average throughput: the total number of tasks that completed their execution per time unit.
- Average turnaround time: the total time required for the process. Starting at task arrival time until the time it finishes execution.

Average number of context switch: The total number of

F. Simulation Results and Analysis

The first evaluation metric is the average throughput. Figure 3 shows the average throughput that is achieved by five algorithms: Hybrid Technique, Neural-based algorithm, RR, WRR and RRR. Figure 3 shows that the average throughput increases as the time passes. Neural Based Algorithm achieves 14% higher throughput improvement than the other three algorithms, and Hybrid Technique achieves 20% higher throughput improvement than the four algorithms, especially when using a time interval greater than 2 Second, when the number of processes is 1000 and the average process duration is 2.5 second.

The second evaluation metric is the average turnaround time. Figure 4 shows the performance of the five algorithms. Neural Based algorithm achieves 18.2% higher improvement than the other three algorithms, and Hybrid technique achieves 22% especially when using a time interval greater than 2.5 Second.

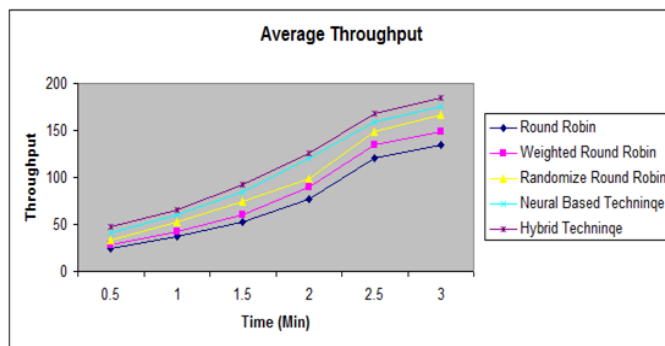


Figure 3. Average throughput.

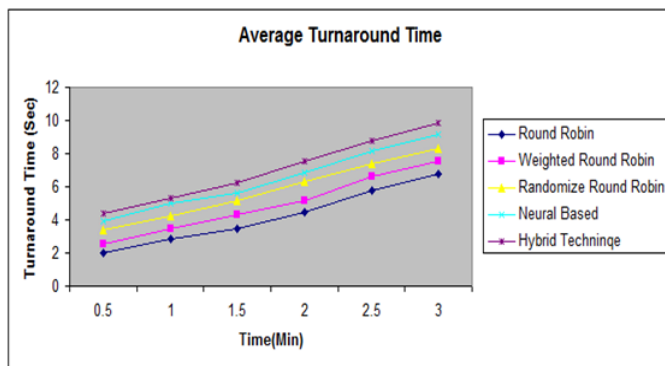


Figure 4. Average turnaround time.

The third evaluation metric is the average context switch, which is presented in Figure 5. As shown in figure 5, Neural Based algorithm achieves the lower number of context switch which lays an improvement of 18% higher than the other four algorithms.

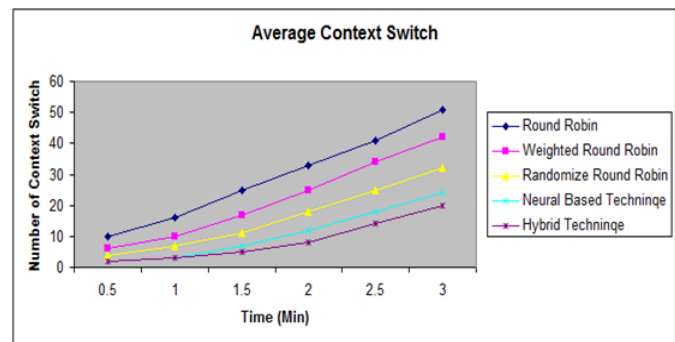


Figure 5. Average context switch.

Furthermore Hybrid Technique achieves the lowest number of context switch which lays an improvement of 20%, especially when using a time interval greater than 2 Second, when the number of processes is 1000 and the average process duration is 2.5 Second.

Fig. 6 depicts the average throughput with different time intervals where; number of processes is 50,000 and the average process duration is 4.6 Seconds. As shown in figure 10; Neural Based Algorithm achieves 18% better performance than the other four algorithms (NB), (RR), (WRR) and (RRR), especially in time 4 second, and hybrid technique achieves 19.5% better performance than the other four algorithms.

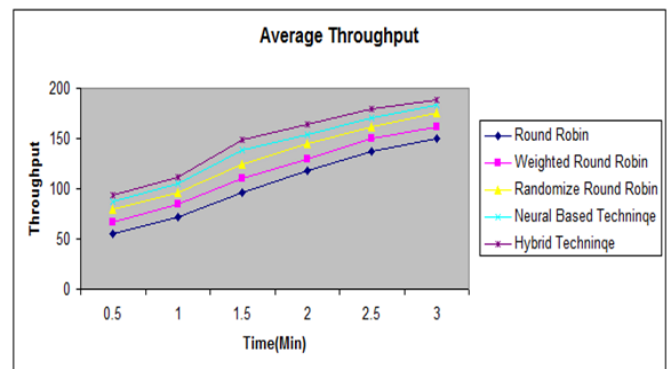


Figure 6. Average throughput.

Fig. 7 depicts the average context switch with different time interval where; number of processes is 50,000 and the average process duration is 4 Seconds. As we show in figure 7; Neural Based Algorithm achieves 18% better performance than the other four algorithms (NB), (RR), (WRR) and (RRR), and Hybrid technique is achieves 23% better performance than the other algorithms.

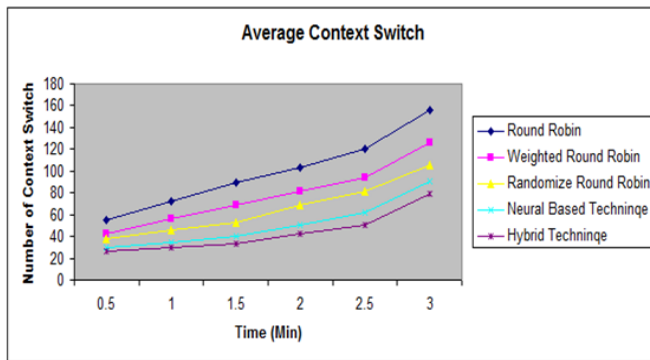


Figure 7. Average context switch.

Fig. 8 depicts the average turnaround time with different time intervals where; number of processes is 50,000 and the average process duration is 4 Seconds. As shown in Figure 8; Neural Based Algorithm achieves 19% higher performance than the other four algorithms (NB), (RR), (WRR) and (RRR), and Hybrid technique achieves 21% higher performance than the other algorithms.

Fig. 9 depicts the average turnaround time with different time intervals where; number of processes is 100,000 and the average process duration is 4 Seconds. As shown in Figure 9; Neural Based Algorithm achieves 19% higher performance than the other four algorithms (NB), (RR), (WRR) and (RRR), and Hybrid technique achieves 21% higher performance than the other algorithms.

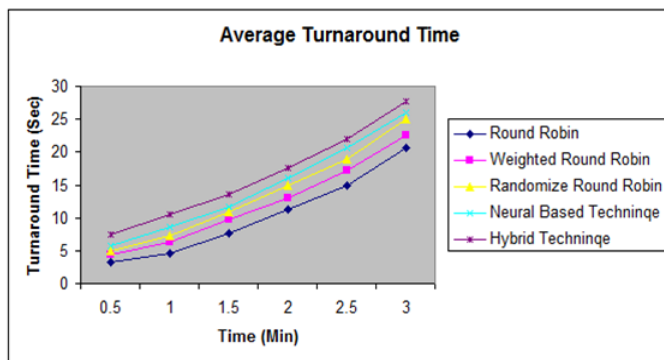


Figure 8. Average turnaround time.

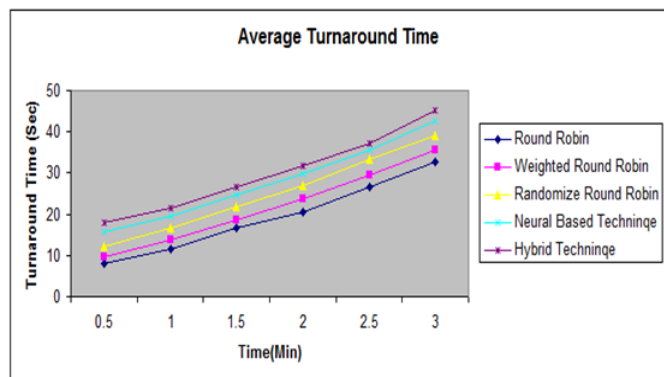


Figure 9. Average turnaround time.

Fig. 10 depicts the average throughput with different time intervals where; number of processes is 100,000 and the average process duration is 4 seconds. As shown in figure 10; Neural Based Algorithm achieves 19% higher performance than the other four algorithms (NB), (RR), (WRR) and (RRR). Hybrid technique achieves 24% higher performance than the other algorithms.

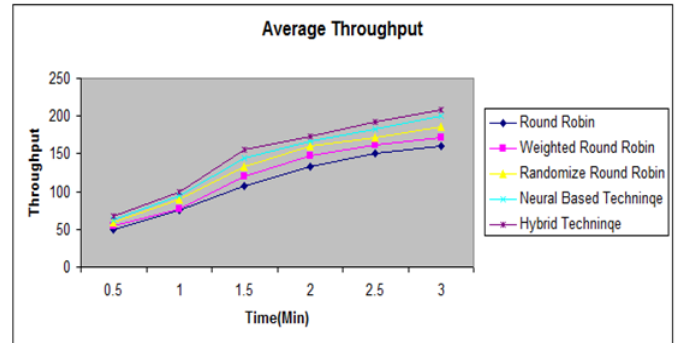


Figure 10. Average throughput.

Fig. 11 depicts the average context switch with different time intervals where; number of processes is 100,000 and the average process duration is 4 Seconds. As shown in figure 11; Neural Based Algorithm achieves 18% better performance than the other four algorithms (NB), (RR), (WRR) and (RRR), Hybrid technique achieves 23% higher performance than the other algorithms.

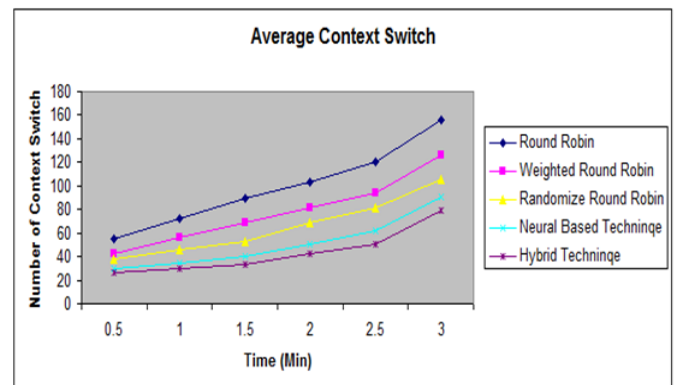


Figure 11. Average context switch.

IV. CONCLUSION AND FUTURE WORKS

Cloud computing is a parallel distributed system that consists many servers that are disturbed in different geographical area, cloud computing consists of virtualized computers and large scale of interconnected based on Service-Level Agreements (SLA) which are established through negotiation between the consumers and the service provider [2]. Furthermore, cloud computing, is concerned with the ability of dynamically providing the amount of required resources to satisfy the customers' demands. Many customers share the resources provided by the cloud, when different users access the cloud data center and request certain jobs, cloud computing must organize and monitor these jobs to achieve fairness among all users. Many job scheduling algorithms have

been proposed to achieve both customer satisfaction and high resource utilization.

In this paper, we proposed a hybrid algorithm for job scheduling in cloud computing system, which based on Neural Network Algorithm. Neural Network distributes the jobs to the four weighted classes depends on job weighs. The experiments showed that the proposed algorithm performed better than RR, WRR, and RRR in terms of average throughput, average turnaround times, and average context switch, which are too important metrics, where the system productivity and the total number of users served per time unit will be more efficient.

The performance of scheduling algorithms actually relies on resource management. Resources are shared between processes based on their class type.

In the future we intend to implement a NN algorithm to learn the usage patterns of resources. Moreover, by predicting such patterns, we can predict when a certain resource will be free. Therefore, we can allow such resources to be borrowed by other processes, which are expected to increase the system performance.

TABLE II. SUMMAIZATION THE EXPERIMENTS RESULTS

Throughput		Turnaround Time		Context switch		Number of Users
Neural Based	Hybrid Technique	Neural Based	Hybrid Technique	Neural Based	Hybrid Technique	
14%	20%	18.2%	22%	18%	20%	1000 Users
18%	19.5%	19%	21%	18%	23%	50000 Users
19%	24%	19%	21%	18%	23%	100000 Users

REFERENCES

- [1] F. Howell, and R. McNab, "SimJava: A discrete event simulation library for java," The first International Conference on Web-Based Modeling and Simulation, 1998.
- [2] M. Arora, K. Das, and R. Biswas, "A Desterilized Scheduling and Load Balancing Algorithm of Heterogeneous Grid Environments," International Conference of Parallel Processing Workshop, 2002, pp. 499-505.
- [3] M. Bani Yassien, Y. Khamayseh, and A. Hatamleh, "Intelligent Randomize Round Robin for Cloud Computing," International journal of cloud applications and computing, 2013, pp. 31-37.
- [4] A. Beitch, B. Liu, T. Yung, R. Grith, and A. Fox, "Rain: A Workload Generation Toolkit for Cloud Computing Applications," University of California, Tech. Rep. UCB/EECS-2010-14, 2010.
- [5] R. Buyya, C. Yeo, S. Venugopal, J. Broberg, and I. Brandic, "Cloud Computing and Emerging IT Platforms: Vision, hype, and reality for delivering computing as the 5th utility," Future Generation Computer Systems, vol. 25, no. 6, pp. 599-616, 2009.
- [6] R. Buyya, R. Ranjan, and R. Calheiros, "Modeling and simulation of scalable cloud computing environments and the CloudSim toolkit. Challenges and opportunities," In Proc. Of (HPCS): The 7th High Performance Computing and Simulation Conference, 2009.
- [7] T. Casavant, and J. Kunl, "A Taxonomy of Scheduling in General Purpose Distributed Computing System," IEEE TRANSACTIONS ON SOFTWARE ENGINEERING, vol. 14, No. 2, pp. 141-154, FEBRUARY 1988.
- [8] D. Chappell, "Introducing the Azure Services Platform," White Paper, sponsored by Microsoft Corporation, 2008.
- [9] X. Fatos, and A. Ajith, "Computational Models and Heuristic Methods for Grid Scheduling Problems," Future Generation Computer System, 2013, pp. 608-621.
- [10] I. Foster, T. Freeman, K. Keahey, D. Scheftner, B. Sotomayor, and X. Zhang, "Virtual clusters for Grid communities," The Sixth Intl. Symposium on Cluster Computing and the Grid, 2006.
- [11] J. Liu, S. Mann, N. V. Vorst, and K. Hellman., "An open and scalable emulation infrastructure for large-scale real-time network simulations," 26th IEEE International Conference on Computer Communications, 2007.
- [12] M. Maheswaran, H. Jay Siegel, D. Hensgen, and R. F. Freund, "Dynamic Mapping of a Class of Independent Tasks onto Heterogeneous Computing Systems," Journal of Parallel and Distributed Computing, vol. 59, pp. 107-131, 1999.
- [13] R. Moreno, and I. Llorente, "Elastic management of cluster-based services in the cloud," 1st Workshop on Automated Control for Datacenters and Clouds (ACDC), held jointly with the Intl. Conference on Autonomic Computing and Communications, Barcelona, Spain, 2009.
- [14] B. Paul, and C. Mark, "Generating Representative Web Workloads for Network and Server Performance Evaluation," ACM SIGMETRICS International Conference on Measurement and Modeling of Computer System, 1998, pp. 151-160.
- [15] P. Saeed, and R. Entezari-Maleki, "RASA, A New Grid Task Scheduling Algorithm," International Journal of Digital Content Technology and its Applications, pp. 91-99, 3 December 2009.
- [16] The Economist, A Special Report on Corporate IT. <http://www.economist.com/specia>, 2008.
- [17] Q. Zhang, L. Cheng, and R. Boutaba, "Cloud computing, state-of-the-art and research challenge," Journal of Internet Services and Applications, 2010, pp. 7-18.
- [18] F. Faniyi, R. Bahsoon, "Engineering Proprioception in SLA Management for Cloud Architectures," Proceedings of the 9th Working IEEE/IFIP Conference on Software Architecture (WICSA), Boulder, CO, USA, 2011, pp. 336-340.
- [19] K. H. Kim, A. Beloglazov, R. Buyya, "Power-aware Provisioning of Cloud Resources for Real-time Services," Proceedings of the 7th international workshop on middleware for Grids, Clouds and e-Science, Urbana Champaign, Illinois, USA, 2009.
- [20] AmazonEC2, Amazon Elastic Compute Cloud (Amazon EC2), <http://aws.amazon.com/ec2/>, 25-4-2013.
- [21] R. Mysore, N. Pamboris, A. Farrington, N. Huang, N. Miri, P. Radhakrishnan, S. Subramanya, V. Vahdat, A. PortLand, "A scalable fault-tolerant layer 2 data center network fabric," SIGCOMM Comput, Commun, Re, 2009, vol. 39, pp. 39-50.
- [22] A. Núñez, J. L. Vázquez-Poletti, A. C. Caminero, G. G. Castañé, J. Carretero, I. M. Llorente, "iCanCloud: A flexible and scalable cloud infrastructure simulator," Journal of Grid Computing, vol. 10, pp. 185-209, Mar 1, 2012.

Residual Energy based One-Hop Data Gathering in wireless sensor networks

Gaurav Kumar Pandey
Dept. of Computer science and engineering
Lovely Professional University
Jalandhar, India

Amritpal Singh
Dept. of Computer science and engineering
Lovely Professional University
Jalandhar, India

Abstract—The key constraint which hampers the performance of Wireless Sensor Networks is the limited battery power of the sensor nodes. Nodes once deployed cannot be recharged therefore data gathering from the sensor field should be done in such a manner that the energy of sensor nodes can be saved. Multi Hop routing and data relay protocols tend to deplete the battery power of the forwarding nodes at a large extent. Also, Clustering Algorithms generate extra overhead which affects the lifetime and performance of the network. In this paper we introduce Residual Energy based One-Hop Data Gathering (REO-HDG) in Wireless Sensor Networks by making use of a Mobile Data Collector (MDC) that traverses the sensor field and collects data from the sensors using single hop only, which in turn eliminates the problem of data relay. We make use of rendezvous locations, one-hop neighbor sets and residual energy of sensors to gather data from the sensor nodes. The union of all neighbor sets include all the candidate sensor nodes. REO-HDG tends to maximize the lifetime of the sensor network by eliminating data relay and clustering.

Index Terms— Mobile Data Collector (MDC), Data gathering, Residual Energy, Energy Conservation, MDC Scheduling, Wireless Sensor Networks.

I. INTRODUCTION

The emergence of Internet of Things (IOT) has attracted a lot of researchers and companies to explore and invest in Wireless Sensor Networks (WSN). WSN has been conducive in mining information from hostile environments where human presence is not possible. Its deployment domains include surveillance, military applications, wildlife monitoring, gathering seismic and volcano outburst related data, under-water habitat monitoring etc. Their impending applications comprises tracking tectonic-plate movements and generating well-timed warnings of natural calamities like Tsunami, Avalanches, and forest fires. But their diminutive battery power has always been a key constraint which hinders their performance. Minimizing the energy dissipation at sensor nodes is a key point for every researcher. Sensor nodes are usually dispersed in a large sensor field which rarely provides any means for recharging the batteries of sensor nodes. So once a sensor node is out of power it cannot be rejuvenated. Keeping this in mind, the operational algorithms of WSN have to be designed in such a manner that the energy level of sensor nodes can be conserved and used judiciously.

The task of WSN includes :- (i) sensing the environment. (ii) Processing the gathered data. (iii) Transferring the gathered data to the base station. Phase of data transfer from the source nodes to the sink node can be achieved by:-

- Multi hop routing or data relay protocols.
- Mobile Elements (ME) or Mobile Data Collectors.

Mobile Data Collector is a robotic agent which is equipped with abundant buffer memory, high capacity rechargeable battery and powerful transceiver. MDC begins its trajectory from a base station, enters the sensor field, gathers data from sensor nodes and finally delivers the accumulated data to the base station. A typical scenario depicting use of MDC is shown in fig.1.

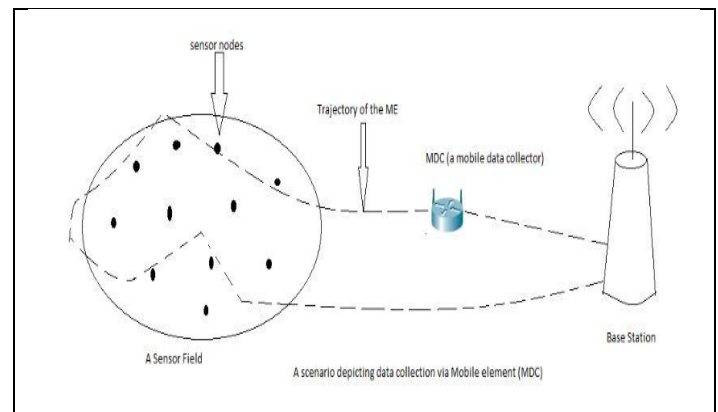


Figure. 1 A scenario depicting data collection via MDC

In our previous work [1] we highlighted the *shortcomings of multi-hop routing protocols* which makes use of intermediate sensor nodes as relay station for transferring the data to the sink. In such type of algorithms the sensor node not only has to sense the environment but also has to act as a forwarding node to relay data of other nearby sensor nodes. This calls for a dense deployment of sensor nodes so that there is end to end connectivity which in turn increases the cost of WSN. Also there are chances of interference due to multi-hop transmission. The *funneling effect* [2] at the nodes near to sink further cause energy depletion and reduces the efficiency. So there are problems

related to *cost*, *connectivity* and *reliability*. These 3 problems can be easily tackled by MDC.

- **Problem related to cost and connectivity-** MEs ameliorates the situation by providing the option of *sparse-deployment* of nodes, since multi-hop transfer is greatly reduced by making use of MDC (i.e. mobile data collectors), MS (mobile sinks) and Re-locatable nodes. There is no need for end to end connectivity since the mobile elements helps to relay or bridge the path between sender and receiver nodes.

- **Problem related to reliability-** Bandwidth is the bone of contention between intermediate nodes during data transfer through multi-hop route. Also there are fading effects and impediments which increases latency and probability of message loss. Mobile elements solve this problem by personally visiting the node that generated the data ->gathering data from that node by single hop transfer mode ->finally takes back the data to the base station or communication end points.

In this paper we introduce *Residual Energy based one-hop mobile data gathering in Wireless Sensor Networks* without making use of any clustering algorithm. According to our algorithm the MDC will visit the *rendezvous locations or probable points* from where the sensor nodes are in the radio range of the MDC or where an actual sensor is physically located. The WSN network is assumed to be sparse and disconnected.

The rest of the paper is organized as follows. Section II gives a survey of related work in the field of mobile data gathering. Section III contains terminologies and assumptions used in the paper. Section IV gives the formulation of Residual Energy based One-Hop Data Gathering (REOHDG) algorithm. Section V contains simulation results and discussions. Finally, Section VI concludes the paper and discusses the future scope.

II. RELATED STUDY

Here we present a brief review of some related works in the field of data gathering in wireless sensor networks.

In our previous work [3] we carried out an extensive survey on recent energy efficient routing protocols in WSN where we reviewed EHCERP [4], BEAR [5], REAMDG [6], SCEP [7], Cluster-based routing using Ant Bee Colony (ABC) algorithm [8] and Virtual Backbone Routing [9].

EHCERP used Balanced Clustering for conserving energy in which proximity from the base station was considered. Cluster heads were categorized into *first level cluster head*, *second level cluster head* and so on depending upon the proximity from base station. Simulation results showed that EHCERP outperformed LEACH, PEGASIS and TEEN. But the major drawback is that re-election of cluster heads cause extra overhead which is a hindrance in maximizing the lifetime of the network.

The technique used by BEAR protocol is *Learning Automata* to discover a compromise between energy balancing and optimal distance. It tends to balance and reduce energy consumption by routing the data according to the residual energy and proximity from the base station. But this protocol too makes use of multi-hop data transfer which reduces the energy of each forwarding node.

REAMDG makes use of a Mobile Collector to gather the data from sensor field which is divided into different clusters using *spectral clustering*. The protocol makes use of residual energy of the nodes and constructs a data relay tree for every cluster. The MDC visits every cluster head to collect data. The algorithm was proved to be *NP Hard*. However the extra overhead caused by creating and managing cluster heads along with data relay tends to affect the overall performance of WSN.

SCEP is an energy efficient protocol which makes use of *map-reduce technique* along with clustering to increase the lifetime of the network. But the main shortcoming is the overhead generated due to the use of key-value pair in the algorithm for the purpose of deciding cluster heads.

Cluster-based routing using *Artificial Bee Colony* algorithm obtains *fitness function* by computing the distance of every node from every cluster head in order to assign sensor node to that cluster head for which there is a minimum distance between node and cluster head. As compared to LEACH, this protocol tends to maximize the lifetime of WSN but yet there is extra overhead involved in calculating fitness function for every node.

Virtual Backbone Routing aims to solve the problem of broadcast storm in WSN by making use of *backbone nodes* which are a subset of the required active nodes that help to minimize and exclude unrequired transmission links by shutting down the radio of extraneous nodes. A schedule transition graph which is a centralized approximation algorithm that maps the backbone node to a particular state.

In our previous work [1], we surveyed mobility based data collection algorithms like *Asynchronous technique* [10], *Scheduled Rendezvous* [11], *Different Energy Radio Technique* [12], *Radio-Triggered Wake-ups* [13], *Stop & Wait Protocol* [14], *Adaptive Speed Control* [15], *Mobi-Route Protocol* [16], *Energy aware Routing to Mobile Gateway* [17], *Two-Tier Data Dissemination Protocol* [18], *Learning Enforced Time Domain Routing* [19], *Network Assisted Navigation* [20], *Mobile Element Scheduling* [21], *Mobile Base station movement using fuzzy logic* [22].

In *Asynchronous technique*, the ME emits signals periodically and the sensing node too wakes up in a periodic fashion to listen and respond to the advertisement signal by initiating data transfer. In case there is no detection of the discovery signal, the sensor node goes back to dormant state. *Scheduled rendezvous* is like a time table system, for example- a bus has a predefined time-table to visit every bus stop on its route at a particular time-slot. These bus stands consists of sensor nodes which become active on arrival of the bus (a MDC) to exchange the data. *Different Energy Radio Technique* makes use of variable energy radio system whereby the ME uses a powerful radio for data communication and a comparatively low range radio for triggering the sensor nodes. *Radio-Triggered Wake-ups* makes use of energy harnessing; MDC sends activation signals to sensor nodes where it is used to activate the transceiver of the sensor node. *Stop & Wait Protocol* made use of a *message loss model* to determine the performance of data collection via MDC. Speeds ranging 30-150 cm/sec, 20-40 km/hour and 1 m/sec were used according to different scenarios. A graph between chance of message loss and distance was

plotted and the resultant curve obtained was found to be parabola. *Adaptive Speed Control* Categorizes the sensor nodes into- LOW, MEDIUM & HIGH groups. The speed of MDC can be increased to $2 \times S$ which will let the MDC complete its journey in $0.5 \times T$. So the remaining time ($T - 0.5T = 0.5T$) can be used by the ME to wait and collect data from sensor nodes efficiently. *Mobi Route Protocol* makes use of time-out signals to determine faulty or terminated links if MDC goes out of the contact area. It increases tolerance towards less optimal routes in order to avoid overhead energy expenditure in re-building the relay tree. Also, it buffers the data in order to reduce loss of data during the movement of MDC. *Energy aware Routing to Mobile Gateway* introduces nodes that can expand their transmission range up to a certain extent to deal with the mobile element which is travelling out of the contact region. The intermediate nodes can be used as forwarders or *relay station* if the MDC is out of coverage range. *Two-Tier Data Dissemination Protocol* involves position-aware routing and builds a grid like structure to forward data to the mobile elements. The forwarding sensor nodes align themselves according to the crossing junctions of the grid network. The intersection point of grid lines denote a crossing point. During the phase of data collection, the mobile elements propagate requests flooding the network. The nodes which are closer to the MDC align themselves on the crossing points and act as disseminating node to spread the request and finally allow the data from the supposed source nodes to pass through them; therefore acting as a *proxy* (a bridge) between MDC and source sensor nodes. In *Learning Enforced Time Domain Routing* the MDC follows a Gaussian mobility distribution pattern. Only those nodes which can interact with MDC via one-hop routes are elected as gateways or proxy nodes. These proxy nodes record the mobility patterns of MDC and deduce/predict a *reinforcement value* which foretells the future arriving of MDC based on probability. Now this value is disseminated to the other nodes so that they can prepare to forward data to the proxy server when the MDC arrives in future. *Network Assisted Navigation* determines a path from which all the nodes can interact via single hop. The authors present the concept of *navigation agents*. These navigation agents are those nodes from which the other sensor nodes can be reached via one-hop route. The authors computed the path along these navigation agents using travelling salesman problem. In *Mobile Element Scheduling*, when the MDC visits the sensor node for data collection, the re-visit time is updated which denotes the time period before which the MDC must visit the sensor node in order to prevent data loss due to buffer overflow at the sensor node. Cost matrix was used for proper scheduling of MDC. *Mobile Base station movement using fuzzy logic* proposed the use of fuzzy logic to manage the movement patterns of the mobile base-station in order to collect data from the static cluster heads such that the energy at the cluster heads is maintained. The authors made use of parameters like: energy of the cluster, size of the cluster and the distance of base station from the cluster head so as to assign a critical degree to the cluster head. Then the tour of base station is devised in such a way that the base station visits those cluster heads which are having a higher value of critical degree.

III. PRELIMINARIES

This paper deals with the problem of data gathering in which the MDC can reach the radio range of every sensor node thereby collecting data via one-hop without making use of any data relay. The sensor network is sparse and partially connected. Also, the sensor nodes are stationary and are visited by the MDC for gathering the sensed data. This section deals with the terminologies and relations that will be used in this paper.

The MDC discussed in this paper is equipped with an Omni-directional antenna that has the same radio range as that of the sensor nodes. All sensor nodes can perform neighbor discovery for their one hop neighbor sensors. We plan an advance trip of MDC to visit the sensor field in order to explore the *rendezvous locations or probable points* where the MDC should stop to collect data from the sensors. These rendezvous locations can be of two categories: (i) coordinate locations where actual physical sensor is located. (ii) Coordinates where MDC has a one hop neighbor sensor i.e. point where MDC is in the radio range of any sensor. As the MDC travels the sensor field, it broadcasts *hello packets* periodically with the radio range similar to that of the sensor nodes. The sensor nodes which are within the range of MDC respond to it via *reply packet*. The reply packet consists of the ID of the sensor, ID of its *one-hop neighbor sensors* (if present) and the *residual energy* of the sensor. The MDC demarks the point where it gets response from the sensor node as a *rendezvous point(r)* and adds the ID of this sensor node into *neighbor set(Nb)* of this rendezvous point. The neighbor set here denotes a set containing the IDs of sensors which were within the wireless range of MDC when it was present at that rendezvous point i.e. one hop neighbor sensors of MDC. If a wireless link exists between a sensor s_i and rendezvous point r_j then sensor s_i belongs to the neighbor set of r_j . Every sensor which has one hop neighbor sensor/s notifies the MDC by piggybacking the IDs of its neighbor sensors in the reply packet. So if at a rendezvous point, the MDC gets a reply packet from a sensor which contains IDs of the neighbors of this sensor, then the MDC moves to the location of this sensor and demarks its location as the next rendezvous point. At this rendezvous point the MDC can collect the data from the nearby neighbor sensors via one hop. So we here present the idea that the MDC will physically visit a sensor node only when this sensor node has some neighbor sensors else it will collect data from the sensor node via one hop from some nearby rendezvous point. The union of all neighbor sets contains all the sensor nodes. As show in fig.2 MDC has reached a point r_0 where it gets a reply packet from s_1 which notifies that it has 2 neighbor sensors s_2 and s_3 . So MDC demarks this point as rendezvous point r_0 and collects data from s_1 . Further MDC puts ID of this sensor in the neighbor set $Nb(r_0)$ and the residual energy of this sensor into $RE[Nb(r_0)]$. Then it moves to the location r_1 which is in fact the location of s_1 . After reaching r_1 , MDC add the nearby sensors in the neighbor set $Nb(r_1)$ and the residual energy of all the neighbor sensors into $RE[Nb(r_1)]$. So in this manner the MDC will locate the rendezvous points and generate neighbor set and residual energy set.

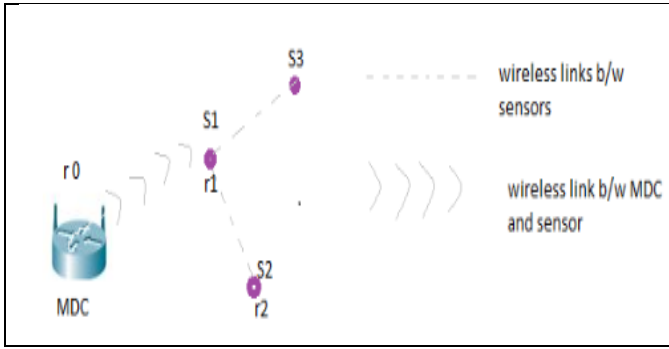


Figure. 2. Locating rendezvous points

IV. HEURISTIC ALGORITHM FOR RESIDUAL ENERGY BASED ONE-HOP DATA GATHERING (REO-HDG)

We assume that every sensor node contains a passive RFID device which can wake up the transceiver of the sensor as soon as it receives a beep from the MDC. In this way the sensors need not to overhear the channel continuously. Also this RFID device doesn't require any sort of power supply, it can get energy from the RF signal of the MDC. Also due to *fading and attenuation* there is a possibility that all the rendezvous points may not be tracked down by the MDC during exploration of the sensor field so we present a heuristic algorithm to solve the problem approximately.

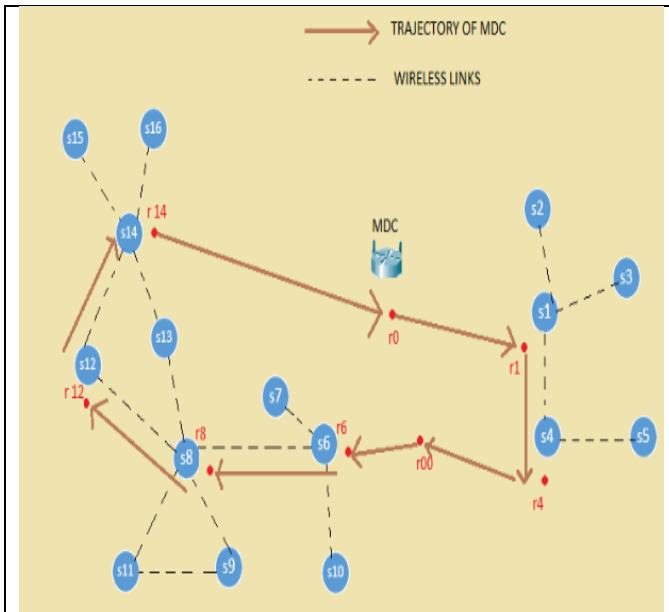


Figure. 3 Residual Energy based One Hop Data Gathering

Figure 3 demonstrates data gathering by REO-HDG algorithm. The above scenario is for the purpose of understanding the working of our algorithm. Given scenario:

Set of sensors (S) = {s1, s2, s3, s4, s5, s6, s7, s8, s9, s10, s11, s12, s13, s14, s15, s16}

Set of rendezvous points (r) = {r0, r1, r4, r00, r6, r8, r12, r14}

Neighbor set (Nb):-

Nb(r0) = {s1}
Nb(r1) = {s2, s3, s4}
Nb(r4) = {s5}
Nb(r00) = {s6}
Nb(r6) = {s7, s10, s8}
Nb(r8) = {s6, s9, s11, s12, s13}
Nb(r12) = {s8, s13, s14}
Nb(r14) = {s12, s15, s16}

Residual Energy set for every neighbor set: RE[Nb]-

RE[Nb(r0)] = min{s1(a)} in mAh ; here a,b,c etc. denote residual energy of the respective sensors.

RE[Nb(r1)] = min{ s2(b), s3(c), s4(d)} in mAh

RE[Nb(r4)] = min{s5(e)} in mAh

RE[Nb(r00)] = min{s6(f)} in mAh

RE[Nb(r6)] = min{s7(g), s10(h), s8(i)} in mAh

RE[Nb(r8)] = min{s6(f), s9(j), s11(k), s12(l), s13(m)}

RE[Nb(r12)] = min{ s8(i), s13(m), s14(n)} in mAh

RE[Nb(r14)] = min{ s12(l), s15(o), s16(p)} in mAh

So we have a set of residual energy containing remaining battery power/energy for every neighbor set. Now during actual data gathering phase the order in which these polling points will be visited by the MDC(mobile data collector) will depend on the MINIMUM VALUE OF RESIDUAL ENERGY i.e. the polling point which has the minimum residual energy (min[R(Nb(11))]). For example:- Nb(r1) has 3 sensors having a, b, c joules/mAh of energy respectively. So calculate min(a,b,c). This will be the residual energy of rendezvous point r1 denoted as RE(r1). And Nb(r2) has 2 sensors having p, q joules of energy respectively. So calculate min(p,q). This will be the residual energy of polling point r2 denoted as RE(r2). NOW if RE(r1) < RE(r2) visit first r1 then r2 and vice-versa.

ALGORITHM FOR PLANNING A PATH OF MDC

- 1) Initially there is an empty set of Path Points (P)
- 2) Generate a set S containing all sensors.
- 3) Create a set r containing all rendezvous points.
- 4) While (S != NULL) // i.e when every sensor has given its data to MDC
 - {
 - Find a rendezvous point (i.e r0,r1,r2...) from r which has minimum value of Residual energy RE.
 - Cover all sensors in Nb(r) i.e. gather data via one hop from neighbor set
 - of above selected polling point and also fetch current residual energy of neighbor sensors and update the set RE[Nb(r)]
 - }

- Add the above selected rendezvous point into Path point (P)
- Now remove the above selected rendezvous point from set r .
- Remove all the covered sensors from set S which are in the neighbor set of the rendezvous point i.e. $Nb(r)$

In fig. 3 the set of Path Points (P) for the shown tour of MDC is $\{r_0, r_1, r_4, r_{00}, r_6, r_8, r_{12}, r_{14}\}$. For planning the next tour the updated set of Residual Energy of rendezvous points will be compared and the minimum residual energy points will be visited first.

V. SIMULATION RESULTS

This section contains the evaluation of REO-HDG algorithm. In our simulation we have assumed that sensor nodes are uniformly distributed in the sensor field and are stationary. For the purpose of comparison we have considered LEACH [23] protocol and SCRC [24] algorithm which we have simulated along with a MDC to gather the data from sensor nodes which are organized into clusters by LEACH and SCRC respectively. We have compared our REOHDG algorithm with these two protocols on the basis of residual energy of the sensor network, number of rounds (MDC tours) and life time of the sensor network. The size of the sensor field is taken to be 500 x 500 m. The transmission range of sensor nodes is taken as 40 m. The MDC starts its trajectory from a particular coordinate (0, 250), covers all the rendezvous points in the increasing order of residual energy and returns back to the same starting point.

A. Energy of sensor network and Number of Rounds

The fig. 4 shows a graph between LEACH and REOHDG in terms of number of rounds and total energy of WSN. Here energy of network represents the initial energy level of sensor network and number of rounds indicate the number of data gathering tours that can be performed. It can be seen that REOHDG outperforms LEACH and can gather the data for a large number of rounds. Hence our algorithm gathers more data from a sensor field and increases the duration for data collection.

B. Sensor Network Lifetime and Residual Energy

The fig.5 shows a graph describing comparison of REOHDG, LEACH and SCRC algorithms based on lifetime and residual energy of the network. It can be seen that residual energy of sensor network decreases rapidly in case of SCRC and LEACH whereas it decreases gradually in case of REOHDG. The tangents obtained in case of LEACH was 0.091 and in case of SCRC was 0.069 while it was 0.056 in case of REOHDG. Clearly our algorithm increases the lifetime of the sensor network up to a major extent because it doesn't involve any extra overheads due to cluster head selection or data relay involved in inter-cluster routing.

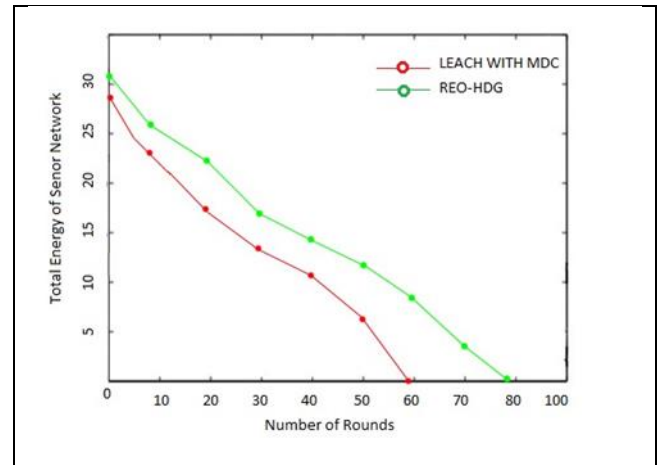


Figure.4 Comparison with LEACH in terms of number of rounds and Total energy of the sensor network

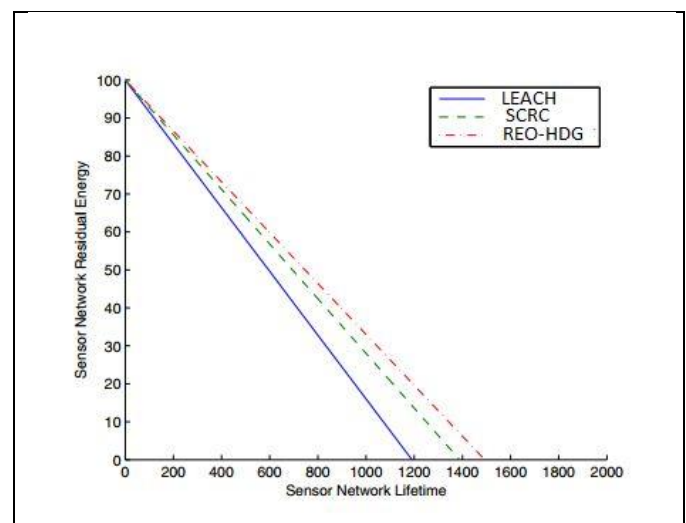


Figure.5 Comparison with SCRC and LEACH in terms of residual energy and lifetime of sensor network

VI. CONCLUSION AND FUTURE WORK

In this paper we have introduced residual energy based data gathering approach without using any clustering or data relay. We made use of a Mobile Data Collector that visits rendezvous locations in order to collect data via one hop. The tour of MDC is based on the residual energy of the rendezvous points. Since the overhead due to cluster head election and multi-hop data forwarding is eliminated therefore the performance is significantly increased. The simulation results show that REO-HDG tends to maximize the lifetime of sensor network as compared to other clustering algorithms. The future scope of this work includes making use of multiple MDCs that can cover separate tours in order to decrease the latency and manage freshness of data in case of very large sensor networks. Also machine learning algorithms like *Reinforcement Learning* can be used whereby the MDC can learn the rendezvous points dynamically thereby making the algorithm adaptive and flexible to unexpected environments.

REFERENCES

- [1] Pandey, Gaurav Kumar, and Amrit Pal Singh. "Energy Conservation and Efficient Data Collection in WSN-ME: A Survey." *Indian Journal of Science and Technology* 8.17 (2015).
- [2] Li J, Mohapatra P. Analytical modeling and mitigation techniques for the energy-hole problem in sensor networks. *Pervasive Mobile Computing*. 2007; 3(3):233–54.
- [3] Pandey, Gaurav Kumar, and Amritpal Singh. "Recent Advancements in Energy Efficient Routing in Wireless Sensor Networks: A Survey." *Proceedings of Fifth International Conference on Soft Computing for Problem Solving*. Springer Singapore, 2016.
- [4] Stefanos A. Nikolidakis, Dionisis Kandris: Energy Efficient Routing in Wireless Sensor Networks through Balanced Clustering. *Algorithms* ISSN 1999-4893 www.mdpi.com / journal /algorithms 6, 29-42; doi:10.3390/a6010029 (2013).
- [5] Ehsan Ahvar, Mahmood Fathy: BEAR: A Balanced Energy-Aware Routing Protocol for Wireless Sensor Networks. *Wireless Sensor Network* 2, 793-800 (2010).
- [6] Rao X, Huang H, Tang J, Zhao H. Residual energy aware mobile data gathering in wireless sensor networks. *Telecommunication Systems*. 2015:1-1.
- [7] Hassan Naderi, Mohammad Reza Kangavari, Morteza Okhovvat: ScEP: A Scalable and Energy Aware Protocol to Increase Network Lifetime in Wireless Sensor Networks, *Wireless Pers Commun*, 82:611–623 © Springer Science+Business Media New York (2015)
- [8] Karaboga, D., & Ozturk, C. A novel clustering approach: Artificial bee colony (ABC) algorithm. *Applied Soft Computing*, 11, 652–657 (2011)
- [9] Umesh B.N, Dr Siddharaju: Energy Efficient Routing of Wireless Sensor Networks Using Virtual Backbone. *International Journal of Wireless & Mobile Networks(IJWMN)*, February(2013)
- [10] Jain, S., Shah, R., Brunette, W., Borriello, G., and Roy, S. 2006. Exploiting mobility for energy efficient data collection in wireless sensor networks. *ACM/Springer Mobile Networks and Applications* 11, 3 (June), 327–339.
- [11] Chakrabarti A, Sabharwal A, Aazhang B. Using predictable observer mobility for power efficient design of sensor networks. *Proceedings of the 2nd International Conference on Information Processing in Sensor Networks (IPSN'03)*; 2003; Volume 2634. p. 129–45.
- [12] Yang X, Vaidya N. A wakeup scheme for sensor networks: Achieving balance between energy saving and end-to-end delay. *2004 Proceedings of 10th IEEE Real-Time and Embedded Technology and Applications (RTAS 2004)*; 2004. p. 19–26.
- [13] Ansari J, Pankin D, Mahonen P. Radio-triggered wake-ups with addressing capabilities for extremely low power sensor network applications. *International Journal of Wireless Information Networks*. 2009 Sep; 16(3):118–30.
- [14] Somasundara A, Kansal A, Jea D, Estrin D, Srivastava M. Controllably mobile infrastructure for low energy embedded networks. *IEEE Transactions on Mobile Computing*. 2006; 5(8):958–73.
- [15] Kansal A, Somasundara A, Jea D, Srivastava M, Estrin D. Intelligent fluid infrastructure for embedded networks. *Proceedings of the 2nd international conference on Mobile systems, applications and services (MobiSys '04)*; 2004. p. 111–24.
- [16] Luo J, Panchard J, Piorkowski M, Grossglauser M, Hubaux JP. 2006. MobiRoute: Routing towards a mobile sink for improving lifetime in sensor networks. *2nd IEEE/ACM DCOSS*. 2006; 4026:480–97
- [17] . Akkaya K, Younis M. Energy-aware routing to a mobile gateway in wireless sensor networks. *IEEE Global Telecommunications Conference Workshops*; 2004. p. 16–21.
- [18] Luo H, Ye F, Cheng J, Lu S, Zhang L. TTDD: two-tier data dissemination in large-scale wireless sensor networks. *Wireless Networks*. 2005; 11(1–2):161–75.
- [19] Baruah P, Urgaonkar R, Krishnamachari B. Learning-enforced time domain routing to mobile sinks in wireless sensor fields. *29th Annual IEEE International Conference on Local Computer Networks*; 2004. p. 525–32.
- [20] Rao J, Biswas S. Network-assisted sync navigation for distributed data gathering - stability and delay - energy trade - offs. *Computer Communications*. 2010; 33(2) - 160 - 75.
- [21] Somasundara AA, Ramamoorthy A, Srivastava MB. Mobile element scheduling for efficient data collection in wireless sensor networks with dynamic deadlines. *2013 IEEE 34th Real-Time Systems Symposium*; 2004. p. 296–305.
- [22] Torghabeh NA, Totonchi MRA. Mobile Base Station Management using Fuzzy Logic in Wireless Sensor Networks. *IEEE 2010 2nd International Conference on Computer Engineering and Technology*; Chengdu; p. V2-352–V26.
- [23] Heinzelman, W.R., Chandrakasan, A., and Balakrishnan, H. (2000). Energy-efficient communication protocol for wireless microsensor networks, In *Proceedings of the 33rd Hawaii International Conference on System Sciences(HICSS-33)*, January (2000).
- [24] Elbhir B, Fkihi SE, Saadane R, Lasaad N, Jorio A, Aboutajdine D. A new spectral classification for robust clustering in wireless sensor networks. In *Wireless and Mobile Networking Conference (WMNC)*, 2013 6th Joint IFIP 2013 Apr 23 (pp. 1-10). IEEE.



Gaurav Kumar Pandey is currently pursuing M.Tech in Computer science and engineering from Lovely Professional University, Jalandhar, India. His specialization stream is computer networking and his research areas include wireless sensor networks, cryptography and routing protocols. He

has authored four research papers targeting energy efficient protocols for sensor networks.



Amritpal Singh has been working as an assistant professor in Lovely Professional University since 2013. He received his M.Tech degree in in Computer science and engineering from Lovely Professional University, Jalandhar, India. He is currently pursuing PhD in Computer Science and

Engineering from Lovely Professional University. His research area includes Database security.

IJCSIS REVIEWERS' LIST

Assist Prof (Dr.) M. Emre Celebi, Louisiana State University in Shreveport, USA
Dr. Lam Hong Lee, Universiti Tunku Abdul Rahman, Malaysia
Dr. Shimon K. Modi, Director of Research BSPA Labs, Purdue University, USA
Dr. Jianguo Ding, Norwegian University of Science and Technology (NTNU), Norway
Assoc. Prof. N. Jaisankar, VIT University, Vellore, Tamilnadu, India
Dr. Amogh Kavimandan, The Mathworks Inc., USA
Dr. Ramasamy Mariappan, Vinayaka Missions University, India
Dr. Yong Li, School of Electronic and Information Engineering, Beijing Jiaotong University, P.R. China
Assist. Prof. Sugam Sharma, NIET, India / Iowa State University, USA
Dr. Jorge A. Ruiz-Vanoye, Universidad Autónoma del Estado de Morelos, Mexico
Dr. Neeraj Kumar, SMVD University, Katra (J&K), India
Dr. Genge Bela, "Petru Maior" University of Targu Mures, Romania
Dr. Junjie Peng, Shanghai University, P. R. China
Dr. Ilhem LENGILIZ, HANA Group - CRISTAL Laboratory, Tunisia
Prof. Dr. Durgesh Kumar Mishra, Acropolis Institute of Technology and Research, Indore, MP, India
Dr. Jorge L. Hernández-Ardieta, University Carlos III of Madrid, Spain
Prof. Dr. C. Suresh Gnana Dhas, Anna University, India
Dr. Li Fang, Nanyang Technological University, Singapore
Prof. Pijush Biswas, RCC Institute of Information Technology, India
Dr. Siddhivinayak Kulkarni, University of Ballarat, Ballarat, Victoria, Australia
Dr. A. Arul Lawrence, Royal College of Engineering & Technology, India
Dr. Wongyos Keardsri, Chulalongkorn University, Bangkok, Thailand
Dr. Somesh Kumar Dewangan, CSVTU Bhilai (C.G.) / Dimat Raipur, India
Dr. Hayder N. Jasem, University Putra Malaysia, Malaysia
Dr. A.V. Senthil Kumar, C. M. S. College of Science and Commerce, India
Dr. R. S. Karthik, C. M. S. College of Science and Commerce, India
Dr. P. Vasant, University Technology Petronas, Malaysia
Dr. Wong Kok Seng, Soongsil University, Seoul, South Korea
Dr. Praveen Ranjan Srivastava, BITS PILANI, India
Dr. Kong Sang Kelvin, Leong, The Hong Kong Polytechnic University, Hong Kong
Dr. Mohd Nazri Ismail, Universiti Kuala Lumpur, Malaysia
Dr. Rami J. Matarneh, Al-isra Private University, Amman, Jordan
Dr. Ojesanmi Olusegun Ayodeji, Ajayi Crowther University, Oyo, Nigeria
Dr. Riktesh Srivastava, Skyline University, UAE
Dr. Oras F. Baker, UCSI University - Kuala Lumpur, Malaysia
Dr. Ahmed S. Ghiduk, Faculty of Science, Beni-Suef University, Egypt
and Department of Computer science, Taif University, Saudi Arabia
Dr. Tirthankar Gayen, IIT Kharagpur, India
Dr. Huei-Ru Tseng, National Chiao Tung University, Taiwan
Prof. Ning Xu, Wuhan University of Technology, China
Dr. Mohammed Salem Binwahlan, Hadhramout University of Science and Technology, Yemen
& Universiti Teknologi Malaysia, Malaysia.
Dr. Aruna Ranganath, Bhoj Reddy Engineering College for Women, India
Dr. Hafeezullah Amin, Institute of Information Technology, KUST, Kohat, Pakistan

Prof. Syed S. Rizvi, University of Bridgeport, USA
Dr. Shahbaz Pervez Chattha, University of Engineering and Technology Taxila, Pakistan
Dr. Shishir Kumar, Jaypee University of Information Technology, Wakanaghat (HP), India
Dr. Shahid Mumtaz, Portugal Telecommunication, Instituto de Telecomunicações (IT) , Aveiro, Portugal
Dr. Rajesh K Shukla, Corporate Institute of Science & Technology Bhopal M P
Dr. Poonam Garg, Institute of Management Technology, India
Dr. S. Mehta, Inha University, Korea
Dr. Dilip Kumar S.M, Bangalore University, Bangalore
Prof. Malik Sikander Hayat Khiyal, Fatima Jinnah Women University, Rawalpindi, Pakistan
Dr. Virendra Gomase , Department of Bioinformatics, Padmashree Dr. D.Y. Patil University
Dr. Irraivan Elamvazuthi, University Technology PETRONAS, Malaysia
Dr. Saqib Saeed, University of Siegen, Germany
Dr. Pavan Kumar Gorakavi, IPMA-USA [YC]
Dr. Ahmed Nabih Zaki Rashed, Menoufia University, Egypt
Prof. Shishir K. Shandilya, Rukmani Devi Institute of Science & Technology, India
Dr. J. Komala Lakshmi, SNR Sons College, Computer Science, India
Dr. Muhammad Sohail, KUST, Pakistan
Dr. Manjaiah D.H, Mangalore University, India
Dr. S Santhosh Baboo, D.G.Vaishnav College, Chennai, India
Prof. Dr. Mokhtar Beldjehem, Sainte-Anne University, Halifax, NS, Canada
Dr. Deepak Laxmi Narasimha, University of Malaya, Malaysia
Prof. Dr. Arunkumar Thangavelu, Vellore Institute Of Technology, India
Dr. M. Azath, Anna University, India
Dr. Md. Rabiul Islam, Rajshahi University of Engineering & Technology (RUET), Bangladesh
Dr. Aos Alaa Zaidan Ansaef, Multimedia University, Malaysia
Dr. Suresh Jain, Devi Ahilya University, Indore (MP) India,
Dr. Mohammed M. Kadhum, Universiti Utara Malaysia
Dr. Hanumanthappa. J. University of Mysore, India
Dr. Syed Ishtiaque Ahmed, Bangladesh University of Engineering and Technology (BUET)
Dr. Akinola Solomon Olalekan, University of Ibadan, Ibadan, Nigeria
Dr. Santosh K. Pandey, The Institute of Chartered Accountants of India
Dr. P. Vasant, Power Control Optimization, Malaysia
Dr. Petr Ivankov, Automatika - S, Russian Federation
Dr. Utkarsh Seetha, Data Infosys Limited, India
Mrs. Priti Maheshwary, Maulana Azad National Institute of Technology, Bhopal
Dr. (Mrs) Padmavathi Ganapathi, Avinashilingam University for Women, Coimbatore
Assist. Prof. A. Neela madheswari, Anna university, India
Prof. Ganesan Ramachandra Rao, PSG College of Arts and Science, India
Mr. Kamanashis Biswas, Daffodil International University, Bangladesh
Dr. Atul Gonsai, Saurashtra University, Gujarat, India
Mr. Angkoon Phinyomark, Prince of Songkla University, Thailand
Mrs. G. Nalini Priya, Anna University, Chennai
Dr. P. Subashini, Avinashilingam University for Women, India
Assoc. Prof. Vijay Kumar Chakka, Dhirubhai Ambani IICT, Gandhinagar ,Gujarat
Mr Jitendra Agrawal, : Rajiv Gandhi Proudhyogiki Vishwavidyalaya, Bhopal
Mr. Vishal Goyal, Department of Computer Science, Punjabi University, India
Dr. R. Baskaran, Department of Computer Science and Engineering, Anna University, Chennai

Assist. Prof. Kanwalvir Singh Dhindsa, B.B.S.B.Engg.College, Fatehgarh Sahib (Punjab), India
Dr. Jamal Ahmad Dargham, School of Engineering and Information Technology, Universiti Malaysia Sabah
Mr. Nitin Bhatia, DAV College, India
Dr. Dhavachelvan Ponnurangam, Pondicherry Central University, India
Dr. Mohd Faizal Abdollah, University of Technical Malaysia, Malaysia
Assist. Prof. Sonal Chawla, Panjab University, India
Dr. Abdul Wahid, AKG Engg. College, Ghaziabad, India
Mr. Arash Habibi Lashkari, University of Malaya (UM), Malaysia
Mr. Md. Rajibul Islam, Ibnu Sina Institute, University Technology Malaysia
Professor Dr. Sabu M. Thampi, .B.S Institute of Technology for Women, Kerala University, India
Mr. Noor Muhammed Nayeem, Université Lumière Lyon 2, 69007 Lyon, France
Dr. Himanshu Aggarwal, Department of Computer Engineering, Punjabi University, India
Prof R. Naidoo, Dept of Mathematics/Center for Advanced Computer Modelling, Durban University of Technology, Durban, South Africa
Prof. Mydhili K Nair, Visweswaraiah Technological University, Bangalore, India
M. Prabu, Adhiyamaan College of Engineering/Anna University, India
Mr. Swakkhar Shatabda, United International University, Bangladesh
Dr. Abdur Rashid Khan, ICIT, Gomal University, Dera Ismail Khan, Pakistan
Mr. H. Abdul Shabeer, I-Nautix Technologies, Chennai, India
Dr. M. Aramudhan, Perunthalaivar Kamarajar Institute of Engineering and Technology, India
Dr. M. P. Thapliyal, Department of Computer Science, HNB Garhwal University (Central University), India
Dr. Shahaboddin Shamshirband, Islamic Azad University, Iran
Mr. Zeashan Hameed Khan, Université de Grenoble, France
Prof. Anil K Ahlawat, Ajay Kumar Garg Engineering College, Ghaziabad, UP Technical University, Lucknow
Mr. Longe Olumide Babatope, University Of Ibadan, Nigeria
Associate Prof. Raman Maini, University College of Engineering, Punjabi University, India
Dr. Maslin Masrom, University Technology Malaysia, Malaysia
Sudipta Chattopadhyay, Jadavpur University, Kolkata, India
Dr. Dang Tuan NGUYEN, University of Information Technology, Vietnam National University - Ho Chi Minh City
Dr. Mary Lourde R., BITS-PILANI Dubai , UAE
Dr. Abdul Aziz, University of Central Punjab, Pakistan
Mr. Karan Singh, Gautam Budtha University, India
Mr. Avinash Pokhriyal, Uttar Pradesh Technical University, Lucknow, India
Associate Prof Dr Zuraini Ismail, University Technology Malaysia, Malaysia
Assistant Prof. Yasser M. Alginahi, Taibah University, Madinah Munawwarah, KSA
Mr. Dakshina Ranjan Kisku, West Bengal University of Technology, India
Mr. Raman Kumar, Dr B R Ambedkar National Institute of Technology, Jalandhar, Punjab, India
Associate Prof. Samir B. Patel, Institute of Technology, Nirma University, India
Dr. M.Munir Ahamed Rabbani, B. S. Abdur Rahman University, India
Asst. Prof. Koushik Majumder, West Bengal University of Technology, India
Dr. Alex Pappachen James, Queensland Micro-nanotechnology center, Griffith University, Australia
Assistant Prof. S. Hariharan, B.S. Abdur Rahman University, India
Asst Prof. Jasmine. K. S, R.V.College of Engineering, India
Mr Naushad Ali Mamode Khan, Ministry of Education and Human Resources, Mauritius
Prof. Mahesh Goyani, G H Patel Collge of Engg. & Tech, V.V.N, Anand, Gujarat, India
Dr. Mana Mohammed, University of Tlemcen, Algeria
Prof. Jatinder Singh, Universal Institution of Engg. & Tech. CHD, India

Mrs. M. Anandhavalli Gauthaman, Sikkim Manipal Institute of Technology, Majitar, East Sikkim
Dr. Bin Guo, Institute Telecom SudParis, France
Mrs. Maleika Mehr Nigar Mohamed Heenaye-Mamode Khan, University of Mauritius
Prof. Pijush Biswas, RCC Institute of Information Technology, India
Mr. V. Bala Dhandayuthapani, Mekelle University, Ethiopia
Dr. Irfan Syamsuddin, State Polytechnic of Ujung Pandang, Indonesia
Mr. Kavi Kumar Khedo, University of Mauritius, Mauritius
Mr. Ravi Chandiran, Zagro Singapore Pte Ltd. Singapore
Mr. Milindkumar V. Sarode, Jawaharlal Darda Institute of Engineering and Technology, India
Dr. Shamimul Qamar, KSJ Institute of Engineering & Technology, India
Dr. C. Arun, Anna University, India
Assist. Prof. M.N.Birje, Basaveshwar Engineering College, India
Prof. Hamid Reza Naji, Department of Computer Enigneering, Shahid Beheshti University, Tehran, Iran
Assist. Prof. Debasis Giri, Department of Computer Science and Engineering, Haldia Institute of Technology
Subhabrata Barman, Haldia Institute of Technology, West Bengal
Mr. M. I. Lali, COMSATS Institute of Information Technology, Islamabad, Pakistan
Dr. Feroz Khan, Central Institute of Medicinal and Aromatic Plants, Lucknow, India
Mr. R. Nagendran, Institute of Technology, Coimbatore, Tamilnadu, India
Mr. Amnach Khawne, King Mongkut's Institute of Technology Ladkrabang, Ladkrabang, Bangkok, Thailand
Dr. P. Chakrabarti, Sir Padampat Singhanian University, Udaipur, India
Mr. Nafiz Imtiaz Bin Hamid, Islamic University of Technology (IUT), Bangladesh.
Shahab-A. Shamshirband, Islamic Azad University, Chalous, Iran
Prof. B. Priestly Shan, Anna Univeristy, Tamilnadu, India
Venkatramreddy Velma, Dept. of Bioinformatics, University of Mississippi Medical Center, Jackson MS USA
Akshi Kumar, Dept. of Computer Engineering, Delhi Technological University, India
Dr. Umesh Kumar Singh, Vikram University, Ujjain, India
Mr. Serguei A. Mokhov, Concordia University, Canada
Mr. Lai Khin Wee, Universiti Teknologi Malaysia, Malaysia
Dr. Awadhesh Kumar Sharma, Madan Mohan Malviya Engineering College, India
Mr. Syed R. Rizvi, Analytical Services & Materials, Inc., USA
Dr. S. Karthik, SNS College of Technology, India
Mr. Syed Qasim Bukhari, CIMET (Universidad de Granada), Spain
Mr. A.D.Potgantwar, Pune University, India
Dr. Himanshu Aggarwal, Punjabi University, India
Mr. Rajesh Ramachandran, Naipunya Institute of Management and Information Technology, India
Dr. K.L. Shunmuganathan, R.M.K Engg College , Kavaraipettai ,Chennai
Dr. Prasant Kumar Pattnaik, KIST, India.
Dr. Ch. Aswani Kumar, VIT University, India
Mr. Ijaz Ali Shoukat, King Saud University, Riyadh KSA
Mr. Arun Kumar, Sir Padam Pat Singhanian University, Udaipur, Rajasthan
Mr. Muhammad Imran Khan, Universiti Teknologi PETRONAS, Malaysia
Dr. Natarajan Meghanathan, Jackson State University, Jackson, MS, USA
Mr. Mohd Zaki Bin Mas'ud, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia
Prof. Dr. R. Geetharamani, Dept. of Computer Science and Eng., Rajalakshmi Engineering College, India
Dr. Smita Rajpal, Institute of Technology and Management, Gurgaon, India
Dr. S. Abdul Khader Jilani, University of Tabuk, Tabuk, Saudi Arabia
Mr. Syed Jamal Haider Zaidi, Bahria University, Pakistan

Dr. N. Devarajan, Government College of Technology, Coimbatore, Tamilnadu, INDIA
Mr. R. Jagadeesh Kannan, RMK Engineering College, India
Mr. Deo Prakash, Shri Mata Vaishno Devi University, India
Mr. Mohammad Abu Naser, Dept. of EEE, IUT, Gazipur, Bangladesh
Assist. Prof. Prasun Ghosal, Bengal Engineering and Science University, India
Mr. Md. Golam Kaosar, School of Engineering and Science, Victoria University, Melbourne City, Australia
Mr. R. Mohammad Shafi, Madanapalle Institute of Technology & Science, India
Dr. F. Sagayaraj Francis, Pondicherry Engineering College, India
Dr. Ajay Goel, HIET, Kaithal, India
Mr. Nayak Sunil Kashibarao, Bahirji Smarak Mahavidyalaya, India
Mr. Suhas J Manangi, Microsoft India
Dr. Kalyankar N. V., Yeshwant Mahavidyalaya, Nanded, India
Dr. K.D. Verma, S.V. College of Post graduate studies & Research, India
Dr. Amjad Rehman, University Technology Malaysia, Malaysia
Mr. Rachit Garg, L K College, Jalandhar, Punjab
Mr. J. William, M.A.M college of Engineering, Trichy, Tamilnadu, India
Prof. Jue-Sam Chou, Nanhua University, College of Science and Technology, Taiwan
Dr. Thorat S.B., Institute of Technology and Management, India
Mr. Ajay Prasad, Sir Padampat Singhanian University, Udaipur, India
Dr. Kamaljit I. Lakhtaria, Atmiya Institute of Technology & Science, India
Mr. Syed Rafiul Hussain, Ahsanullah University of Science and Technology, Bangladesh
Mrs Fazeela Tunnisa, Najran University, Kingdom of Saudi Arabia
Mrs Kavita Taneja, Maharishi Markandeshwar University, Haryana, India
Mr. Maniyar Shiraz Ahmed, Najran University, Najran, KSA
Mr. Anand Kumar, AMC Engineering College, Bangalore
Dr. Rakesh Chandra Gangwar, Beant College of Engg. & Tech., Gurdaspur (Punjab) India
Dr. V V Rama Prasad, Sree Vidyanikethan Engineering College, India
Assist. Prof. Neetesh Kumar Gupta, Technocrats Institute of Technology, Bhopal (M.P.), India
Mr. Ashish Seth, Uttar Pradesh Technical University, Lucknow, UP India
Dr. V V S S S Balaram, Sreenidhi Institute of Science and Technology, India
Mr Rahul Bhatia, Lingaya's Institute of Management and Technology, India
Prof. Niranjana Reddy, P, KITS, Warangal, India
Prof. Rakesh. Lingappa, Vijetha Institute of Technology, Bangalore, India
Dr. Mohammed Ali Hussain, Nimra College of Engineering & Technology, Vijayawada, A.P., India
Dr. A. Srinivasan, MNM Jain Engineering College, Rajiv Gandhi Salai, Thorapakkam, Chennai
Mr. Rakesh Kumar, M.M. University, Mullana, Ambala, India
Dr. Lena Khaled, Zarqa Private University, Aman, Jordan
Ms. Supriya Kapoor, Patni/Lingaya's Institute of Management and Tech., India
Dr. Tossapon Boongoen, Aberystwyth University, UK
Dr. Bilal Alatas, Firat University, Turkey
Assist. Prof. Jyoti Praakash Singh, Academy of Technology, India
Dr. Ritu Soni, GNG College, India
Dr. Mahendra Kumar, Sagar Institute of Research & Technology, Bhopal, India.
Dr. Binod Kumar, Lakshmi Narayan College of Tech. (LNCT) Bhopal India
Dr. Muzhir Shaban Al-Ani, Amman Arab University Amman – Jordan
Dr. T.C. Manjunath, ATRIA Institute of Tech, India
Mr. Muhammad Zakarya, COMSATS Institute of Information Technology (CIIT), Pakistan

Assist. Prof. Harmunish Taneja, M. M. University, India
Dr. Chitra Dhawale , SICSR, Model Colony, Pune, India
Mrs Sankari Muthukaruppan, Nehru Institute of Engineering and Technology, Anna University, India
Mr. Aaqif Afzaal Abbasi, National University Of Sciences And Technology, Islamabad
Prof. Ashutosh Kumar Dubey, Trinity Institute of Technology and Research Bhopal, India
Mr. G. Appasami, Dr. Pauls Engineering College, India
Mr. M Yasin, National University of Science and Tech, karachi (NUST), Pakistan
Mr. Yaser Miaji, University Utara Malaysia, Malaysia
Mr. Shah Ahsanul Haque, International Islamic University Chittagong (IIUC), Bangladesh
Prof. (Dr) Syed Abdul Sattar, Royal Institute of Technology & Science, India
Dr. S. Sasikumar, Roever Engineering College
Assist. Prof. Monit Kapoor, Maharishi Markandeshwar University, India
Mr. Nwaocha Vivian O, National Open University of Nigeria
Dr. M. S. Vijaya, GR Govindarajulu School of Applied Computer Technology, India
Assist. Prof. Chakresh Kumar, Manav Rachna International University, India
Mr. Kunal Chadha , R&D Software Engineer, Gemalto, Singapore
Mr. Mueen Uddin, Universiti Teknologi Malaysia, UTM , Malaysia
Dr. Dhuha Basheer abdullah, Mosul university, Iraq
Mr. S. Audithan, Annamalai University, India
Prof. Vijay K Chaudhari, Technocrats Institute of Technology , India
Associate Prof. Mohd Ilyas Khan, Technocrats Institute of Technology , India
Dr. Vu Thanh Nguyen, University of Information Technology, HoChiMinh City, VietNam
Assist. Prof. Anand Sharma, MITS, Lakshmangarh, Sikar, Rajasthan, India
Prof. T V Narayana Rao, HITAM Engineering college, Hyderabad
Mr. Deepak Gour, Sir Padampat Singhania University, India
Assist. Prof. Amutharaj Joyson, Kalasalingam University, India
Mr. Ali Balador, Islamic Azad University, Iran
Mr. Mohit Jain, Maharaja Surajmal Institute of Technology, India
Mr. Dilip Kumar Sharma, GLA Institute of Technology & Management, India
Dr. Debojyoti Mitra, Sir padampat Singhania University, India
Dr. Ali Dehghantanha, Asia-Pacific University College of Technology and Innovation, Malaysia
Mr. Zhao Zhang, City University of Hong Kong, China
Prof. S.P. Setty, A.U. College of Engineering, India
Prof. Patel Rakeshkumar Kantilal, Sankalchand Patel College of Engineering, India
Mr. Biswajit Bhowmik, Bengal College of Engineering & Technology, India
Mr. Manoj Gupta, Apex Institute of Engineering & Technology, India
Assist. Prof. Ajay Sharma, Raj Kumar Goel Institute Of Technology, India
Assist. Prof. Ramveer Singh, Raj Kumar Goel Institute of Technology, India
Dr. Hanan Elazhary, Electronics Research Institute, Egypt
Dr. Hosam I. Faiq, USM, Malaysia
Prof. Dipti D. Patil, MAEER's MIT College of Engg. & Tech, Pune, India
Assist. Prof. Devendra Chack, BCT Kumaon engineering College Dwarahat Almora, India
Prof. Manpreet Singh, M. M. Engg. College, M. M. University, India
Assist. Prof. M. Sadiq ali Khan, University of Karachi, Pakistan
Mr. Prasad S. Halgaonkar, MIT - College of Engineering, Pune, India
Dr. Imran Ghani, Universiti Teknologi Malaysia, Malaysia
Prof. Varun Kumar Kakar, Kumaon Engineering College, Dwarahat, India

Assist. Prof. Nisheeth Joshi, Apaji Institute, Banasthali University, Rajasthan, India
Associate Prof. Kunwar S. Vaisla, VCT Kumaon Engineering College, India
Prof Anupam Choudhary, Bhilai School Of Engg., Bhilai (C.G.), India
Mr. Divya Prakash Shrivastava, Al Jabal Al garbi University, Zawya, Libya
Associate Prof. Dr. V. Radha, Avinashilingam Deemed university for women, Coimbatore.
Dr. Kasarapu Ramani, JNT University, Anantapur, India
Dr. Anuraag Awasthi, Jayoti Vidyapeeth Womens University, India
Dr. C G Ravichandran, R V S College of Engineering and Technology, India
Dr. Mohamed A. Deriche, King Fahd University of Petroleum and Minerals, Saudi Arabia
Mr. Abbas Karimi, Universiti Putra Malaysia, Malaysia
Mr. Amit Kumar, Jaypee University of Engg. and Tech., India
Dr. Nikolai Stoianov, Defense Institute, Bulgaria
Assist. Prof. S. Ranichandra, KSR College of Arts and Science, Tiruchencode
Mr. T.K.P. Rajagopal, Diamond Horse International Pvt Ltd, India
Dr. Md. Ekramul Hamid, Rajshahi University, Bangladesh
Mr. Hemanta Kumar Kalita, TATA Consultancy Services (TCS), India
Dr. Messaouda Azzouzi, Ziane Achour University of Djelfa, Algeria
Prof. (Dr.) Juan Jose Martinez Castillo, "Gran Mariscal de Ayacucho" University and Acantelys research Group, Venezuela
Dr. Jatinderkumar R. Saini, Narmada College of Computer Application, India
Dr. Babak Bashari Rad, University Technology of Malaysia, Malaysia
Dr. Nighat Mir, Effat University, Saudi Arabia
Prof. (Dr.) G.M.Nasira, Sasurie College of Engineering, India
Mr. Varun Mittal, Gemalto Pte Ltd, Singapore
Assist. Prof. Mrs P. Banumathi, Kathir College Of Engineering, Coimbatore
Assist. Prof. Quan Yuan, University of Wisconsin-Stevens Point, US
Dr. Pranam Paul, Narula Institute of Technology, Agarpara, West Bengal, India
Assist. Prof. J. Ramkumar, V.L.B Janakiammal college of Arts & Science, India
Mr. P. Sivakumar, Anna university, Chennai, India
Mr. Md. Humayun Kabir Biswas, King Khalid University, Kingdom of Saudi Arabia
Mr. Mayank Singh, J.P. Institute of Engg & Technology, Meerut, India
HJ. Kamaruzaman Jusoff, Universiti Putra Malaysia
Mr. Nikhil Patrick Lobo, CADES, India
Dr. Amit Wason, Rayat-Bahra Institute of Engineering & Boi-Technology, India
Dr. Rajesh Shrivastava, Govt. Benazir Science & Commerce College, Bhopal, India
Assist. Prof. Vishal Bharti, DCE, Gurgaon
Mrs. Sunita Bansal, Birla Institute of Technology & Science, India
Dr. R. Sudhakar, Dr.Mahalingam college of Engineering and Technology, India
Dr. Amit Kumar Garg, Shri Mata Vaishno Devi University, Katra(J&K), India
Assist. Prof. Raj Gaurang Tiwari, AZAD Institute of Engineering and Technology, India
Mr. Hamed Taherdoost, Tehran, Iran
Mr. Amin Daneshmand Malayeri, YRC, IAU, Malayer Branch, Iran
Mr. Shantanu Pal, University of Calcutta, India
Dr. Terry H. Walcott, E-Promag Consultancy Group, United Kingdom
Dr. Ezekiel U OKIKE, University of Ibadan, Nigeria
Mr. P. Mahalingam, Caledonian College of Engineering, Oman
Dr. Mahmoud M. A. Abd Ellatif, Mansoura University, Egypt

Prof. Kunwar S. Vaisla, BCT Kumaon Engineering College, India
Prof. Mahesh H. Panchal, Kalol Institute of Technology & Research Centre, India
Mr. Muhammad Asad, Technical University of Munich, Germany
Mr. AliReza Shams Shafigh, Azad Islamic university, Iran
Prof. S. V. Nagaraj, RMK Engineering College, India
Mr. Ashikali M Hasan, Senior Researcher, CelNet security, India
Dr. Adnan Shahid Khan, University Technology Malaysia, Malaysia
Mr. Prakash Gajanan Burade, Nagpur University/ITM college of engg, Nagpur, India
Dr. Jagdish B.Helonde, Nagpur University/ITM college of engg, Nagpur, India
Professor, Doctor BOUHORMA Mohammed, Univertsity Abdelmalek Essaadi, Morocco
Mr. K. Thirumalaivasan, Pondicherry Engg. College, India
Mr. Umbarkar Anantkumar Janardan, Walchand College of Engineering, India
Mr. Ashish Chaurasia, Gyan Ganga Institute of Technology & Sciences, India
Mr. Sunil Taneja, Kurukshetra University, India
Mr. Fauzi Adi Rafrastara, Dian Nuswantoro University, Indonesia
Dr. Yaduvir Singh, Thapar University, India
Dr. Ioannis V. Koskosas, University of Western Macedonia, Greece
Dr. Vasantha Kalyani David, Avinashilingam University for women, Coimbatore
Dr. Ahmed Mansour Manasrah, Universiti Sains Malaysia, Malaysia
Miss. Nazanin Sadat Kazazi, University Technology Malaysia, Malaysia
Mr. Saeed Rasouli Heikalabad, Islamic Azad University - Tabriz Branch, Iran
Assoc. Prof. Dharendra Mishra, SVKM's NMIMS University, India
Prof. Shapoor Zarei, UAE Inventors Association, UAE
Prof. B.Raja Sarath Kumar, Lenora College of Engineering, India
Dr. Bashir Alam, Jamia millia Islamia, Delhi, India
Prof. Anant J Umbarkar, Walchand College of Engg., India
Assist. Prof. B. Bharathi, Sathyabama University, India
Dr. Fokrul Alom Mazarbhuiya, King Khalid University, Saudi Arabia
Prof. T.S.Jeyali Laseeth, Anna University of Technology, Tirunelveli, India
Dr. M. Balraju, Jawahar Lal Nehru Technological University Hyderabad, India
Dr. Vijayalakshmi M. N., R.V.College of Engineering, Bangalore
Prof. Walid Moudani, Lebanese University, Lebanon
Dr. Saurabh Pal, VBS Purvanchal University, Jaunpur, India
Associate Prof. Suneet Chaudhary, Dehradun Institute of Technology, India
Associate Prof. Dr. Manuj Darbari, BBD University, India
Ms. Prema Selvaraj, K.S.R College of Arts and Science, India
Assist. Prof. Ms.S.Sasikala, KSR College of Arts & Science, India
Mr. Sukhvinder Singh Deora, NC Institute of Computer Sciences, India
Dr. Abhay Bansal, Amity School of Engineering & Technology, India
Ms. Sumita Mishra, Amity School of Engineering and Technology, India
Professor S. Viswanadha Raju, JNT University Hyderabad, India
Mr. Asghar Shahrzad Khashandarag, Islamic Azad University Tabriz Branch, India
Mr. Manoj Sharma, Panipat Institute of Engg. & Technology, India
Mr. Shakeel Ahmed, King Faisal University, Saudi Arabia
Dr. Mohamed Ali Mahjoub, Institute of Engineer of Monastir, Tunisia
Mr. Adri Jovin J.J., SriGuru Institute of Technology, India
Dr. Sukumar Senthilkumar, Universiti Sains Malaysia, Malaysia

Mr. Rakesh Bharati, Dehradun Institute of Technology Dehradun, India
Mr. Shervan Fekri Ershad, Shiraz International University, Iran
Mr. Md. Safiqul Islam, Daffodil International University, Bangladesh
Mr. Mahmudul Hasan, Daffodil International University, Bangladesh
Prof. Mandakini Tayade, UIT, RGTU, Bhopal, India
Ms. Sarla More, UIT, RGTU, Bhopal, India
Mr. Tushar Hrishikesh Jaware, R.C. Patel Institute of Technology, Shirpur, India
Ms. C. Divya, Dr G R Damodaran College of Science, Coimbatore, India
Mr. Fahimuddin Shaik, Annamacharya Institute of Technology & Sciences, India
Dr. M. N. Giri Prasad, JNTUCE, Pulivendula, A.P., India
Assist. Prof. Chintan M Bhatt, Charotar University of Science And Technology, India
Prof. Sahista Machchhar, Marwadi Education Foundation's Group of institutions, India
Assist. Prof. Navnish Goel, S. D. College Of Engineering & Technology, India
Mr. Khaja Kamaluddin, Sirt University, Sirt, Libya
Mr. Mohammad Zaidul Karim, Daffodil International, Bangladesh
Mr. M. Vijayakumar, KSR College of Engineering, Tiruchengode, India
Mr. S. A. Ahsan Rajon, Khulna University, Bangladesh
Dr. Muhammad Mohsin Nazir, LCW University Lahore, Pakistan
Mr. Mohammad Asadul Hoque, University of Alabama, USA
Mr. P.V.Sarathchand, Indur Institute of Engineering and Technology, India
Mr. Durgesh Samadhiya, Chung Hua University, Taiwan
Dr Venu Kuthadi, University of Johannesburg, Johannesburg, RSA
Dr. (Er) Jasvir Singh, Guru Nanak Dev University, Amritsar, Punjab, India
Mr. Jasmin Cosic, Min. of the Interior of Una-sana canton, B&H, Bosnia and Herzegovina
Dr S. Rajalakshmi, Botho College, South Africa
Dr. Mohamed Sarrab, De Montfort University, UK
Mr. Basappa B. Kodada, Canara Engineering College, India
Assist. Prof. K. Ramana, Annamacharya Institute of Technology and Sciences, India
Dr. Ashu Gupta, Apeejay Institute of Management, Jalandhar, India
Assist. Prof. Shaik Rasool, Shadan College of Engineering & Technology, India
Assist. Prof. K. Suresh, Annamacharya Institute of Tech & Sci. Rajampet, AP, India
Dr . G. Singaravel, K.S.R. College of Engineering, India
Dr B. G. Geetha, K.S.R. College of Engineering, India
Assist. Prof. Kavita Choudhary, ITM University, Gurgaon
Dr. Mehrdad Jalali, Azad University, Mashhad, Iran
Megha Goel, Shamli Institute of Engineering and Technology, Shamli, India
Mr. Chi-Hua Chen, Institute of Information Management, National Chiao-Tung University, Taiwan (R.O.C.)
Assoc. Prof. A. Rajendran, RVS College of Engineering and Technology, India
Assist. Prof. S. Jaganathan, RVS College of Engineering and Technology, India
Assoc. Prof. (Dr.) A S N Chakravarthy, JNTUK University College of Engineering Vizianagaram (State University)
Assist. Prof. Deepshikha Patel, Technocrat Institute of Technology, India
Assist. Prof. Maram Balajee, GMRIT, India
Assist. Prof. Monika Bhatnagar, TIT, India
Prof. Gaurang Panchal, Charotar University of Science & Technology, India
Prof. Anand K. Tripathi, Computer Society of India
Prof. Jyoti Chaudhary, High Performance Computing Research Lab, India
Assist. Prof. Supriya Raheja, ITM University, India

Dr. Pankaj Gupta, Microsoft Corporation, U.S.A.
Assist. Prof. Panchamukesh Chandaka, Hyderabad Institute of Tech. & Management, India
Prof. Mohan H.S, SJB Institute Of Technology, India
Mr. Hossein Malekinezhad, Islamic Azad University, Iran
Mr. Zatin Gupta, Universti Malaysia, Malaysia
Assist. Prof. Amit Chauhan, Phonics Group of Institutions, India
Assist. Prof. Ajal A. J., METS School Of Engineering, India
Mrs. Omowunmi Omobola Adeyemo, University of Ibadan, Nigeria
Dr. Bharat Bhushan Agarwal, I.F.T.M. University, India
Md. Nazrul Islam, University of Western Ontario, Canada
Tushar Kanti, L.N.C.T, Bhopal, India
Er. Aumreesh Kumar Saxena, SIRTs College Bhopal, India
Mr. Mohammad Monirul Islam, Daffodil International University, Bangladesh
Dr. Kashif Nisar, University Utara Malaysia, Malaysia
Dr. Wei Zheng, Rutgers Univ/ A10 Networks, USA
Associate Prof. Rituraj Jain, Vyas Institute of Engg & Tech, Jodhpur – Rajasthan
Assist. Prof. Apoorvi Sood, I.T.M. University, India
Dr. Kayhan Zrar Ghafoor, University Technology Malaysia, Malaysia
Mr. Swapnil Sonar, Truba Institute College of Engineering & Technology, Indore, India
Ms. Yogita Gigras, I.T.M. University, India
Associate Prof. Neelima Sadineni, Pydha Engineering College, India Pydha Engineering College
Assist. Prof. K. Deepika Rani, HITAM, Hyderabad
Ms. Shikha Maheshwari, Jaipur Engineering College & Research Centre, India
Prof. Dr V S Giridhar Akula, Avanthi's Scientific Tech. & Research Academy, Hyderabad
Prof. Dr.S.Saravanan, Muthayammal Engineering College, India
Mr. Mehdi Golsorkhatabar Amiri, Islamic Azad University, Iran
Prof. Amit Sadanand Savyanavar, MITCOE, Pune, India
Assist. Prof. P.Oliver Jayaprakash, Anna University, Chennai
Assist. Prof. Ms. Sujata, ITM University, Gurgaon, India
Dr. Asoke Nath, St. Xavier's College, India
Mr. Masoud Rafighi, Islamic Azad University, Iran
Assist. Prof. RamBabu Pemula, NIMRA College of Engineering & Technology, India
Assist. Prof. Ms Rita Chhikara, ITM University, Gurgaon, India
Mr. Sandeep Maan, Government Post Graduate College, India
Prof. Dr. S. Muralidharan, Mepco Schlenk Engineering College, India
Associate Prof. T.V.Sai Krishna, QIS College of Engineering and Technology, India
Mr. R. Balu, Bharathiar University, Coimbatore, India
Assist. Prof. Shekhar. R, Dr.SM College of Engineering, India
Prof. P. Senthilkumar, Vivekanandha Institue of Engineering and Technology for Woman, India
Mr. M. Kamarajan, PSNA College of Engineering & Technology, India
Dr. Angajala Srinivasa Rao, Jawaharlal Nehru Technical University, India
Assist. Prof. C. Venkatesh, A.I.T.S, Rajampet, India
Mr. Afshin Rezakhani Roozbahani, Ayatollah Boroujerdi University, Iran
Mr. Laxmi chand, SCTL, Noida, India
Dr. Dr. Abdul Hannan, Vivekanand College, Aurangabad
Prof. Mahesh Panchal, KITRC, Gujarat
Dr. A. Subramani, K.S.R. College of Engineering, Tiruchengode

Assist. Prof. Prakash M, Rajalakshmi Engineering College, Chennai, India
Assist. Prof. Akhilesh K Sharma, Sir Padampat Singhanian University, India
Ms. Varsha Sahni, Guru Nanak Dev Engineering College, Ludhiana, India
Associate Prof. Trilochan Rout, NM Institute of Engineering and Technology, India
Mr. Srikanta Kumar Mohapatra, NMIET, Orissa, India
Mr. Waqas Haider Bangyal, Iqra University Islamabad, Pakistan
Dr. S. Vijayaragavan, Christ College of Engineering and Technology, Pondicherry, India
Prof. Elbouchari Mohamed, University Mohammed First, Oujda, Morocco
Dr. Muhammad Asif Khan, King Faisal University, Saudi Arabia
Dr. Nagy Ramadan Darwish Omran, Cairo University, Egypt.
Assistant Prof. Anand Nayyar, KCL Institute of Management and Technology, India
Mr. G. Premsankar, Ericsson, India
Assist. Prof. T. Hemalatha, VELS University, India
Prof. Tejaswini Apte, University of Pune, India
Dr. Edmund Ng Giap Weng, Universiti Malaysia Sarawak, Malaysia
Mr. Mahdi Nouri, Iran University of Science and Technology, Iran
Associate Prof. S. Asif Hussain, Annamacharya Institute of technology & Sciences, India
Mrs. Kavita Pabreja, Maharaja Surajmal Institute (an affiliate of GGSIP University), India
Mr. Vorugunti Chandra Sekhar, DA-IICT, India
Mr. Muhammad Najmi Ahmad Zabidi, Universiti Teknologi Malaysia, Malaysia
Dr. Aderemi A. Atayero, Covenant University, Nigeria
Assist. Prof. Osama Sohaib, Balochistan University of Information Technology, Pakistan
Assist. Prof. K. Suresh, Annamacharya Institute of Technology and Sciences, India
Mr. Hassen Mohammed Abdullaah Alsafi, International Islamic University Malaysia (IIUM) Malaysia
Mr. Robail Yasrab, Virtual University of Pakistan, Pakistan
Mr. R. Balu, Bharathiar University, Coimbatore, India
Prof. Anand Nayyar, KCL Institute of Management and Technology, Jalandhar
Assoc. Prof. Vivek S Deshpande, MIT College of Engineering, India
Prof. K. Saravanan, Anna university Coimbatore, India
Dr. Ravendra Singh, MJP Rohilkhand University, Bareilly, India
Mr. V. Mathivanan, IBRA College of Technology, Sultanate of OMAN
Assoc. Prof. S. Asif Hussain, AITS, India
Assist. Prof. C. Venkatesh, AITS, India
Mr. Sami Ulhaq, SZABIST Islamabad, Pakistan
Dr. B. Justus Rabi, Institute of Science & Technology, India
Mr. Anuj Kumar Yadav, Dehradun Institute of technology, India
Mr. Alejandro Mosquera, University of Alicante, Spain
Assist. Prof. Arjun Singh, Sir Padampat Singhanian University (SPSU), Udaipur, India
Dr. Smriti Agrawal, JB Institute of Engineering and Technology, Hyderabad
Assist. Prof. Swathi Sambangi, Visakha Institute of Engineering and Technology, India
Ms. Prabhjot Kaur, Guru Gobind Singh Indraprastha University, India
Mrs. Samaher AL-Hothali, Yanbu University College, Saudi Arabia
Prof. Rajneeshkaur Bedi, MIT College of Engineering, Pune, India
Mr. Hassen Mohammed Abdullaah Alsafi, International Islamic University Malaysia (IIUM)
Dr. Wei Zhang, Amazon.com, Seattle, WA, USA
Mr. B. Santhosh Kumar, C S I College of Engineering, Tamil Nadu
Dr. K. Reji Kumar, N S S College, Pandalam, India

Assoc. Prof. K. Seshadri Sastry, EILM University, India
Mr. Kai Pan, UNC Charlotte, USA
Mr. Ruikar Sachin, SGGSIET, India
Prof. (Dr.) Vinodani Katiyar, Sri Ramswaroop Memorial University, India
Assoc. Prof., M. Giri, Sreenivasa Institute of Technology and Management Studies, India
Assoc. Prof. Labib Francis Gergis, Misr Academy for Engineering and Technology (MET), Egypt
Assist. Prof. Amanpreet Kaur, ITM University, India
Assist. Prof. Anand Singh Rajawat, Shri Vaishnav Institute of Technology & Science, Indore
Mrs. Hadeel Saleh Haj Aliwi, Universiti Sains Malaysia (USM), Malaysia
Dr. Abhay Bansal, Amity University, India
Dr. Mohammad A. Mezher, Fahad Bin Sultan University, KSA
Assist. Prof. Nidhi Arora, M.C.A. Institute, India
Prof. Dr. P. Suresh, Karpagam College of Engineering, Coimbatore, India
Dr. Kannan Balasubramanian, Mepco Schlenk Engineering College, India
Dr. S. Sankara Gomathi, Panimalar Engineering college, India
Prof. Anil kumar Suthar, Gujarat Technological University, L.C. Institute of Technology, India
Assist. Prof. R. Hubert Rajan, NOORUL ISLAM UNIVERSITY, India
Assist. Prof. Dr. Jyoti Mahajan, College of Engineering & Technology
Assist. Prof. Homam Reda El-Taj, College of Network Engineering, Saudi Arabia & Malaysia
Mr. Bijan Paul, Shahjalal University of Science & Technology, Bangladesh
Assoc. Prof. Dr. Ch V Phani Krishna, KL University, India
Dr. Vishal Bhatnagar, Ambedkar Institute of Advanced Communication Technologies & Research, India
Dr. Lamri LAOUAMER, Al Qassim University, Dept. Info. Systems & European University of Brittany, Dept. Computer Science, UBO, Brest, France
Prof. Ashish Babanrao Sasankar, G.H.Raisoni Institute Of Information Technology, India
Prof. Pawan Kumar Goel, Shamli Institute of Engineering and Technology, India
Mr. Ram Kumar Singh, S.V Subharti University, India
Assistant Prof. Sunish Kumar O S, Amalijothei College of Engineering, India
Dr Sanjay Bhargava, Banasthali University, India
Mr. Pankaj S. Kulkarni, AVEW's Shatabdi Institute of Technology, India
Mr. Roohollah Etemadi, Islamic Azad University, Iran
Mr. Oloruntoyin Sefiu Taiwo, Emmanuel Alayande College Of Education, Nigeria
Mr. Sumit Goyal, National Dairy Research Institute, India
Mr Jaswinder Singh Dilawari, Geeta Engineering College, India
Prof. Raghuraj Singh, Harcourt Butler Technological Institute, Kanpur
Dr. S.K. Mahendran, Anna University, Chennai, India
Dr. Amit Wason, Hindustan Institute of Technology & Management, Punjab
Dr. Ashu Gupta, Apeejay Institute of Management, India
Assist. Prof. D. Asir Antony Gnana Singh, M.I.E.T Engineering College, India
Mrs Mina Farmanbar, Eastern Mediterranean University, Famagusta, North Cyprus
Mr. Maram Balajee, GMR Institute of Technology, India
Mr. Moiz S. Ansari, Isra University, Hyderabad, Pakistan
Mr. Adebayo, Olawale Surajudeen, Federal University of Technology Minna, Nigeria
Mr. Jasvir Singh, University College Of Engg., India
Mr. Vivek Tiwari, MANIT, Bhopal, India
Assoc. Prof. R. Navaneethakrishnan, Bharathiyar College of Engineering and Technology, India
Mr. Somdip Dey, St. Xavier's College, Kolkata, India

Mr. Souleymane Balla-Arabé, Xi'an University of Electronic Science and Technology, China
Mr. Mahabub Alam, Rajshahi University of Engineering and Technology, Bangladesh
Mr. Sathyapraksh P., S.K.P Engineering College, India
Dr. N. Karthikeyan, SNS College of Engineering, Anna University, India
Dr. Binod Kumar, JSPM's, Jayawant Technical Campus, Pune, India
Assoc. Prof. Dinesh Goyal, Suresh Gyan Vihar University, India
Mr. Md. Abdul Ahad, K L University, India
Mr. Vikas Bajpai, The LNM IIT, India
Dr. Manish Kumar Anand, Salesforce (R & D Analytics), San Francisco, USA
Assist. Prof. Dheeraj Murari, Kumaon Engineering College, India
Assoc. Prof. Dr. A. Muthukumaravel, VELS University, Chennai
Mr. A. Siles Balasingh, St.Joseph University in Tanzania, Tanzania
Mr. Ravindra Daga Badgujar, R C Patel Institute of Technology, India
Dr. Preeti Khanna, SVKM's NMIMS, School of Business Management, India
Mr. Kumar Dayanand, Cambridge Institute of Technology, India
Dr. Syed Asif Ali, SMI University Karachi, Pakistan
Prof. Pallvi Pandit, Himachal Pradesh University, India
Mr. Ricardo Verschueren, University of Gloucestershire, UK
Assist. Prof. Mamta Juneja, University Institute of Engineering and Technology, Panjab University, India
Assoc. Prof. P. Surendra Varma, NRI Institute of Technology, JNTU Kakinada, India
Assist. Prof. Gaurav Shrivastava, RGPV / SVITS Indore, India
Dr. S. Sumathi, Anna University, India
Assist. Prof. Ankita M. Kapadia, Charotar University of Science and Technology, India
Mr. Deepak Kumar, Indian Institute of Technology (BHU), India
Dr. Dr. Rajan Gupta, GGSIP University, New Delhi, India
Assist. Prof M. Anand Kumar, Karpagam University, Coimbatore, India
Mr. Mr Arshad Mansoor, Pakistan Aeronautical Complex
Mr. Kapil Kumar Gupta, Ansal Institute of Technology and Management, India
Dr. Neeraj Tomer, SINE International Institute of Technology, Jaipur, India
Assist. Prof. Trunal J. Patel, C.G.Patel Institute of Technology, Uka Tarsadia University, Bardoli, Surat
Mr. Sivakumar, Codework solutions, India
Mr. Mohammad Sadegh Mirzaei, PGNR Company, Iran
Dr. Gerard G. Dumancas, Oklahoma Medical Research Foundation, USA
Mr. Varadala Sridhar, Varadhaman College Engineering College, Affiliated To JNTU, Hyderabad
Assist. Prof. Manoj Dhawan, SVITS, Indore
Assoc. Prof. Chitreshh Banerjee, Suresh Gyan Vihar University, Jaipur, India
Dr. S. Santhi, SCSVMV University, India
Mr. Davood Mohammadi Souran, Ministry of Energy of Iran, Iran
Mr. Shamim Ahmed, Bangladesh University of Business and Technology, Bangladesh
Mr. Sandeep Reddivari, Mississippi State University, USA
Assoc. Prof. Ousmane Thiare, Gaston Berger University, Senegal
Dr. Hazra Imran, Athabasca University, Canada
Dr. Setu Kumar Chaturvedi, Technocrats Institute of Technology, Bhopal, India
Mr. Mohd Dilshad Ansari, Jaypee University of Information Technology, India
Ms. Jaspreet Kaur, Distance Education LPU, India
Dr. D. Nagarajan, Salalah College of Technology, Sultanate of Oman
Dr. K.V.N.R.Sai Krishna, S.V.R.M. College, India

Mr. Himanshu Pareek, Center for Development of Advanced Computing (CDAC), India
Mr. Khaldi Amine, Badji Mokhtar University, Algeria
Mr. Mohammad Sadegh Mirzaei, Scientific Applied University, Iran
Assist. Prof. Khyati Chaudhary, Ram-eesh Institute of Engg. & Technology, India
Mr. Sanjay Agal, Pacific College of Engineering Udaipur, India
Mr. Abdul Mateen Ansari, King Khalid University, Saudi Arabia
Dr. H.S. Behera, Veer Surendra Sai University of Technology (VSSUT), India
Dr. Shrikant Tiwari, Shri Shankaracharya Group of Institutions (SSGI), India
Prof. Ganesh B. Regulwar, Shri Shankarprasad Agnihotri College of Engg, India
Prof. Pinnamaneni Bhanu Prasad, Matrix vision GmbH, Germany
Dr. Shrikant Tiwari, Shri Shankaracharya Technical Campus (SSTC), India
Dr. Siddesh G.K., : Dayananada Sagar College of Engineering, Bangalore, India
Dr. Nadir Bouchama, CERIST Research Center, Algeria
Dr. R. Sathishkumar, Sri Venkateswara College of Engineering, India
Assistant Prof (Dr.) Mohamed Moussaoui, Abdelmalek Essaadi University, Morocco
Dr. S. Malathi, Panimalar Engineering College, Chennai, India
Dr. V. Subedha, Panimalar Institute of Technology, Chennai, India
Dr. Prashant Panse, Swami Vivekanand College of Engineering, Indore, India
Dr. Hamza Aldabbas, Al-Balqa'a Applied University, Jordan
Dr. G. Rasitha Banu, Vel's University, Chennai
Dr. V. D. Ambeth Kumar, Panimalar Engineering College, Chennai
Prof. Anuranjan Misra, Bhagwant Institute of Technology, Ghaziabad, India
Ms. U. Sinthuja, PSG college of arts & science, India
Dr. Ehsan Saradar Torshizi, Urmia University, Iran
Dr. Shamneesh Sharma, APG Shimla University, Shimla (H.P.), India
Assistant Prof. A. S. Syed Navaz, Muthayammal College of Arts & Science, India
Assistant Prof. Ranjit Panigrahi, Sikkim Manipal Institute of Technology, Majitar, Sikkim
Dr. Khaled Eskaf, Arab Academy for Science ,Technology & Maritime Transportation, Egypt
Dr. Nishant Gupta, University of Jammu, India
Assistant Prof. Nagarajan Sankaran, Annamalai University, Chidambaram, Tamilnadu, India
Assistant Prof. Tribikram Pradhan, Manipal Institute of Technology, India
Dr. Nasser Lotfi, Eastern Mediterranean University, Northern Cyprus
Dr. R. Manavalan, K S Rangasamy college of Arts and Science, Tamilnadu, India
Assistant Prof. P. Krishna Sankar, K S Rangasamy college of Arts and Science, Tamilnadu, India
Dr. Rahul Malik, Cisco Systems, USA
Dr. S. C. Lingareddy, ALPHA College of Engineering, India
Assistant Prof. Mohammed Shuaib, Interat University, Lucknow, India
Dr. Sachin Yele, Sanghvi Institute of Management & Science, India
Dr. T. Thambidurai, Sun Univercell, Singapore
Prof. Anandkumar Telang, BKIT, India
Assistant Prof. R. Poorvadevi, SCSVMV University, India
Dr Uttam Mande, Gitam University, India
Dr. Poornima Girish Naik, Shahu Institute of Business Education and Research (SIBER), India
Prof. Md. Abu Kausar, Jaipur National University, Jaipur, India
Dr. Mohammed Zuber, AISECT University, India
Prof. Kalum Priyanath Udagepola, King Abdulaziz University, Saudi Arabia
Dr. K. R. Ananth, Velalar College of Engineering and Technology, India

Assistant Prof. Sanjay Sharma, Roorkee Engineering & Management Institute Shamli (U.P), India
Assistant Prof. Panem Charan Arur, Priyadarshini Institute of Technology, India
Dr. Ashwak Mahmood muhsen alabaichi, Karbala University / College of Science, Iraq
Dr. Urmila Shrawankar, G H Raison College of Engineering, Nagpur (MS), India
Dr. Krishan Kumar Paliwal, Panipat Institute of Engineering & Technology, India
Dr. Mukesh Negi, Tech Mahindra, India
Dr. Anuj Kumar Singh, Amity University Gurgaon, India
Dr. Babar Shah, Gyeongsang National University, South Korea
Assistant Prof. Jayprakash Upadhyay, SRI-TECH Jabalpur, India
Assistant Prof. Varadala Sridhar, Vidya Jyothi Institute of Technology, India
Assistant Prof. Parameshachari B D, KSIT, Bangalore, India
Assistant Prof. Ankit Garg, Amity University, Haryana, India
Assistant Prof. Rajashe Karappa, SDMCET, Karnataka, India
Assistant Prof. Varun Jasuja, GNIT, India
Assistant Prof. Sonal Honale, Abha Gaikwad Patil College of Engineering Nagpur, India
Dr. Pooja Choudhary, CT Group of Institutions, NIT Jalandhar, India
Dr. Faouzi Hidoussi, UHL Batna, Algeria
Dr. Naseer Ali Hussein, Wasit University, Iraq
Assistant Prof. Vinod Kumar Shukla, Amity University, Dubai
Dr. Ahmed Farouk Metwaly, K L University
Mr. Mohammed Noaman Murad, Cihan University, Iraq
Dr. Suxing Liu, Arkansas State University, USA
Dr. M. Gomathi, Velalar College of Engineering and Technology, India
Assistant Prof. Sumardiono, College PGRI Blitar, Indonesia
Dr. Latika Kharb, Jagan Institute of Management Studies (JIMS), Delhi, India
Associate Prof. S. Raja, Pauls College of Engineering and Technology, Tamilnadu, India
Assistant Prof. Seyed Reza Pakize, Shahid Sani High School, Iran
Dr. Thiyagu Nagaraj, University-INOUE, India
Assistant Prof. Noreen Sarai, Harare Institute of Technology, Zimbabwe
Assistant Prof. Gajanand Sharma, Suresh Gyan Vihar University Jaipur, Rajasthan, India
Assistant Prof. Mapari Vikas Prakash, Siddhant COE, Sudumbare, Pune, India
Dr. Devesh Katiyar, Shri Ramswaroop Memorial University, India
Dr. Shenshen Liang, University of California, Santa Cruz, US
Assistant Prof. Mohammad Abu Omar, Limkokwing University of Creative Technology- Malaysia
Mr. Snehasis Banerjee, Tata Consultancy Services, India
Assistant Prof. Kibona Lusekelo, Ruaha Catholic University (RUCU), Tanzania
Assistant Prof. Adib Kabir Chowdhury, University College Technology Sarawak, Malaysia
Dr. Ying Yang, Computer Science Department, Yale University, USA
Dr. Vinay Shukla, Institute Of Technology & Management, India
Dr. Liviu Octavian Maftciu-Scai, West University of Timisoara, Romania
Assistant Prof. Rana Khudhair Abbas Ahmed, Al-Rafidain University College, Iraq
Assistant Prof. Nitin A. Naik, S.R.T.M. University, India
Dr. Timothy Powers, University of Hertfordshire, UK
Dr. S. Prasath, Bharathiar University, Erode, India
Dr. Ritu Shrivastava, SIRTIS Bhopal, India
Prof. Rohit Shrivastava, Mittal Institute of Technology, Bhopal, India
Dr. Gianina Mihai, Dunarea de Jos" University of Galati, Romania

Assistant Prof. Ms. T. Kalai Selvi, Erode Sengunthar Engineering College, India
Assistant Prof. Ms. C. Kavitha, Erode Sengunthar Engineering College, India
Assistant Prof. K. Sinivasamoorthi, Erode Sengunthar Engineering College, India
Assistant Prof. Mallikarjun C Sarsamba Bheemna Khandre Institute Technology, Bhalki, India
Assistant Prof. Vishwanath Chikaraddi, Veermata Jijabai technological Institute (Central Technological Institute), India
Assistant Prof. Dr. Ikvinderpal Singh, Trai Shatabdi GGS Khalsa College, India
Assistant Prof. Mohammed Noaman Murad, Cihan University, Iraq
Professor Yousef Farhaoui, Moulay Ismail University, Errachidia, Morocco
Dr. Parul Verma, Amity University, India
Professor Yousef Farhaoui, Moulay Ismail University, Errachidia, Morocco
Assistant Prof. Madhavi Dhingra, Amity University, Madhya Pradesh, India
Assistant Prof.. G. Selvavinayagam, SNS College of Technology, Coimbatore, India
Assistant Prof. Madhavi Dhingra, Amity University, MP, India
Professor Kartheesan Log, Anna University, Chennai
Professor Vasudeva Acharya, Shri Madhwa vadiraja Institute of Technology, India
Dr. Asif Iqbal Hajamydeen, Management & Science University, Malaysia
Assistant Prof., Mahendra Singh Meena, Amity University Haryana
Assistant Professor Manjeet Kaur, Amity University Haryana
Dr. Mohamed Abd El-Basset Matwalli, Zagazig University, Egypt
Dr. Ramani Kannan, Universiti Teknologi PETRONAS, Malaysia
Assistant Prof. S. Jagadeesan Subramaniam, Anna University, India
Assistant Prof. Dharmendra Choudhary, Tripura University, India
Assistant Prof. Deepika Vodnala, SR Engineering College, India
Dr. Kai Cong, Intel Corporation & Computer Science Department, Portland State University, USA
Dr. Kailas R Patil, Vishwakarma Institute of Information Technology (VIIT), India
Dr. Omar A. Alzubi, Faculty of IT / Al-Balqa Applied University, Jordan
Assistant Prof. Kareemullah Shaik, Nimra Institute of Science and Technology, India
Assistant Prof. Chirag Modi, NIT Goa
Dr. R. Ramkumar, Nandha Arts And Science College, India
Dr. Priyadarshini Vydhialingam, Harathiar University, India
Dr. P. S. Jagadeesh Kumar, DBIT, Bangalore, Karnataka
Dr. Vikas Thada, AMITY University, Pachgaon
Dr. T. A. Ashok Kumar, Institute of Management, Christ University, Bangalore
Dr. Shaheera Rashwan, Informatics Research Institute
Dr. S. Preetha Gunasekar, Bharathiyar University, India
Asst Professor Sameer Dev Sharma, Uttaranchal University, Dehradun
Dr. Zhihan Iv, Chinese Academy of Science, China
Dr. Ikvinderpal Singh, Trai Shatabdi GGS Khalsa College, Amritsar
Dr. Umar Ruhi, University of Ottawa, Canada
Dr. Jasmin Cosic, University of Bihac, Bosnia and Herzegovina
Dr. Homam Reda El-Taj, University of Tabuk, Kingdom of Saudi Arabia
Dr. Mostafa Ghobaei Arani, Islamic Azad University, Iran
Dr. Ayyasamy Ayyanar, Annamalai University, India
Dr. Selvakumar Manickam, Universiti Sains Malaysia, Malaysia
Dr. Murali Krishna Namana, GITAM University, India
Dr. Smriti Agrawal, Chaitanya Bharathi Institute of Technology, Hyderabad, India
Professor Vimalathithan Rathinasabapathy, Karpagam College Of Engineering, India

Dr. Sushil Chandra Dimri, Graphic Era University, India
Dr. Dinh-Sinh Mai, Le Quy Don Technical University, Vietnam
Dr. S. Rama Sree, Aditya Engg. College, India
Dr. Ehab T. Alnfwawy, Sadat Academy, Egypt
Dr. Patrick D. Cerna, Haramaya University, Ethiopia
Dr. Vishal Jain, Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), India
Associate Prof. Dr. Jiliang Zhang, North Eastern University, China
Dr. Sharefa Murad, Middle East University, Jordan
Dr. Ajeet Singh Poonia, Govt. College of Engineering & technology, Rajasthan, India
Dr. Vahid Esmaeelzadeh, University of Science and Technology, Iran
Dr. Jacek M. Czerniak, Casimir the Great University in Bydgoszcz, Institute of Technology, Poland
Associate Prof. Anisur Rehman Nasir, Jamia Millia Islamia University
Assistant Prof. Imran Ahmad, COMSATS Institute of Information Technology, Pakistan
Professor Ghulam Qasim, Preston University, Islamabad, Pakistan
Dr. Parameshachari B D, GSSS Institute of Engineering and Technology for Women
Dr. Wencan Luo, University of Pittsburgh, US
Dr. Musa PEKER, Faculty of Technology, Mugla Sitki Kocman University, Turkey
Dr. Gunasekaran Shanmugam, Anna University, India
Dr. Binh P. Nguyen, National University of Singapore, Singapore
Dr. Rajkumar Jain, Indian Institute of Technology Indore, India
Dr. Imtiaz Ali Halepoto, QUEST Nawabshah, Pakistan
Dr. Shaligram Prajapat, Devi Ahilya University Indore India
Dr. Sunita Singhal, Birla Institute of Technology and Science, Pilani, India
Dr. Ijaz Ali Shoukat, King Saud University, Saudi Arabia
Dr. Anuj Gupta, IKG Punjab Technical University, India
Dr. Sonali Saini, IES-IPS Academy, India
Dr. Krishan Kumar, Moti Lal Nehru National Institute of Technology, Allahabad, India
Dr. Z. Faizal Khan, College of Engineering, Shaqra University, Kingdom of Saudi Arabia
Prof. M. Padmavathamma, S.V. University Tirupati, India
Prof. A. Velayudham, Cape Institute of Technology, India
Prof. Seifeidne Kadry, American University of the Middle East
Dr. J. Durga Prasad Rao, Pt. Ravishankar Shukla University, Raipur
Assistant Prof. Najam Hasan, Dhofar University
Dr. G. Suseendran, Vels University, Pallavaram, Chennai
Prof. Ankit Faldu, Gujarat Technological University- Atmiya Institute of Technology and Science
Dr. Ali Habiboghli, Islamic Azad University
Dr. Deepak Dembla, JECRC University, Jaipur, India
Dr. Pankaj Rajan, Walmart Labs, USA
Assistant Prof. Radoslava Kraveva, South-West University "Neofit Rilski", Bulgaria
Assistant Prof. Medhavi Shriwas, Shri vaishnav institute of Technology, India
Associate Prof. Sedat Akleylek, Ondokuz Mayıs University, Turkey
Dr. U.V. Arivazhagu, Kingston Engineering College Affiliated To Anna University, India
Dr. Touseef Ali, University of Engineering and Technology, Taxila, Pakistan
Assistant Prof. Naren Jeeva, SASTRA University, India
Dr. Riccardo Colella, University of Salento, Italy
Dr. Enache Maria Cristina, University of Galati, Romania
Dr. Senthil P, Kurinji College of Arts & Science, India

CALL FOR PAPERS

International Journal of Computer Science and Information Security

IJCSIS 2016

ISSN: 1947-5500

<http://sites.google.com/site/ijcsis/>

International Journal Computer Science and Information Security, IJCSIS, is the premier scholarly venue in the areas of computer science and security issues. IJCSIS 2011 will provide a high profile, leading edge platform for researchers and engineers alike to publish state-of-the-art research in the respective fields of information technology and communication security. The journal will feature a diverse mixture of publication articles including core and applied computer science related topics.

Authors are solicited to contribute to the special issue by submitting articles that illustrate research results, projects, surveying works and industrial experiences that describe significant advances in the following areas, but are not limited to. Submissions may span a broad range of topics, e.g.:

Track A: Security

Access control, Anonymity, Audit and audit reduction & Authentication and authorization, Applied cryptography, Cryptanalysis, Digital Signatures, Biometric security, Boundary control devices, Certification and accreditation, Cross-layer design for security, Security & Network Management, Data and system integrity, Database security, Defensive information warfare, Denial of service protection, Intrusion Detection, Anti-malware, Distributed systems security, Electronic commerce, E-mail security, Spam, Phishing, E-mail fraud, Virus, worms, Trojan Protection, Grid security, Information hiding and watermarking & Information survivability, Insider threat protection, Integrity

Intellectual property protection, Internet/Intranet Security, Key management and key recovery, Language-based security, Mobile and wireless security, Mobile, Ad Hoc and Sensor Network Security, Monitoring and surveillance, Multimedia security ,Operating system security, Peer-to-peer security, Performance Evaluations of Protocols & Security Application, Privacy and data protection, Product evaluation criteria and compliance, Risk evaluation and security certification, Risk/vulnerability assessment, Security & Network Management, Security Models & protocols, Security threats & countermeasures (DDoS, MiM, Session Hijacking, Replay attack etc.), Trusted computing, Ubiquitous Computing Security, Virtualization security, VoIP security, Web 2.0 security, Submission Procedures, Active Defense Systems, Adaptive Defense Systems, Benchmark, Analysis and Evaluation of Security Systems, Distributed Access Control and Trust Management, Distributed Attack Systems and Mechanisms, Distributed Intrusion Detection/Prevention Systems, Denial-of-Service Attacks and Countermeasures, High Performance Security Systems, Identity Management and Authentication, Implementation, Deployment and Management of Security Systems, Intelligent Defense Systems, Internet and Network Forensics, Large-scale Attacks and Defense, RFID Security and Privacy, Security Architectures in Distributed Network Systems, Security for Critical Infrastructures, Security for P2P systems and Grid Systems, Security in E-Commerce, Security and Privacy in Wireless Networks, Secure Mobile Agents and Mobile Code, Security Protocols, Security Simulation and Tools, Security Theory and Tools, Standards and Assurance Methods, Trusted Computing, Viruses, Worms, and Other Malicious Code, World Wide Web Security, Novel and emerging secure architecture, Study of attack strategies, attack modeling, Case studies and analysis of actual attacks, Continuity of Operations during an attack, Key management, Trust management, Intrusion detection techniques, Intrusion response, alarm management, and correlation analysis, Study of tradeoffs between security and system performance, Intrusion tolerance systems, Secure protocols, Security in wireless networks (e.g. mesh networks, sensor networks, etc.), Cryptography and Secure Communications, Computer Forensics, Recovery and Healing, Security Visualization, Formal Methods in Security, Principles for Designing a Secure Computing System, Autonomic Security, Internet Security, Security in Health Care Systems, Security Solutions Using Reconfigurable Computing, Adaptive and Intelligent Defense Systems, Authentication and Access control, Denial of service attacks and countermeasures, Identity, Route and

Location Anonymity schemes, Intrusion detection and prevention techniques, Cryptography, encryption algorithms and Key management schemes, Secure routing schemes, Secure neighbor discovery and localization, Trust establishment and maintenance, Confidentiality and data integrity, Security architectures, deployments and solutions, Emerging threats to cloud-based services, Security model for new services, Cloud-aware web service security, Information hiding in Cloud Computing, Securing distributed data storage in cloud, Security, privacy and trust in mobile computing systems and applications, **Middleware security & Security features:** middleware software is an asset on

its own and has to be protected, interaction between security-specific and other middleware features, e.g., context-awareness, **Middleware-level security monitoring and measurement:** metrics and mechanisms for quantification and evaluation of security enforced by the middleware, **Security co-design:** trade-off and co-design between application-based and middleware-based security, **Policy-based management:** innovative support for policy-based definition and enforcement of security concerns, **Identification and authentication mechanisms:** Means to capture application specific constraints in defining and enforcing access control rules, **Middleware-oriented security patterns:** identification of patterns for sound, reusable security, **Security in aspect-based middleware:** mechanisms for isolating and enforcing security aspects, **Security in agent-based platforms:** protection for mobile code and platforms, Smart Devices: Biometrics, National ID cards, Embedded Systems Security and TPMs, RFID Systems Security, Smart Card Security, Pervasive Systems: Digital Rights Management (DRM) in pervasive environments, Intrusion Detection and Information Filtering, Localization Systems Security (Tracking of People and Goods), Mobile Commerce Security, Privacy Enhancing Technologies, Security Protocols (for Identification and Authentication, Confidentiality and Privacy, and Integrity), Ubiquitous Networks: Ad Hoc Networks Security, Delay-Tolerant Network Security, Domestic Network Security, Peer-to-Peer Networks Security, Security Issues in Mobile and Ubiquitous Networks, Security of GSM/GPRS/UMTS Systems, Sensor Networks Security, Vehicular Network Security, Wireless Communication Security: Bluetooth, NFC, WiFi, WiMAX, WiMedia, others

This Track will emphasize the design, implementation, management and applications of computer communications, networks and services. Topics of mostly theoretical nature are also welcome, provided there is clear practical potential in applying the results of such work.

Track B: Computer Science

Broadband wireless technologies: LTE, WiMAX, WiRAN, HSDPA, HSUPA, Resource allocation and interference management, Quality of service and scheduling methods, Capacity planning and dimensioning, Cross-layer design and Physical layer based issue, Interworking architecture and interoperability, Relay assisted and cooperative communications, Location and provisioning and mobility management, Call admission and flow/congestion control, Performance optimization, Channel capacity modeling and analysis, Middleware Issues: Event-based, publish/subscribe, and message-oriented middleware, Reconfigurable, adaptable, and reflective middleware approaches, Middleware solutions for reliability, fault tolerance, and quality-of-service, Scalability of middleware, Context-aware middleware, Autonomic and self-managing middleware, Evaluation techniques for middleware solutions, Formal methods and tools for designing, verifying, and evaluating, middleware, Software engineering techniques for middleware, Service oriented middleware, Agent-based middleware, Security middleware, Network Applications: Network-based automation, Cloud applications, Ubiquitous and pervasive applications, Collaborative applications, RFID and sensor network applications, Mobile applications, Smart home applications, Infrastructure monitoring and control applications, Remote health monitoring, GPS and location-based applications, Networked vehicles applications, Alert applications, Embedded Computer System, Advanced Control Systems, and Intelligent Control : Advanced control and measurement, computer and microprocessor-based control, signal processing, estimation and identification techniques, application specific IC's, nonlinear and adaptive control, optimal and robot control, intelligent control, evolutionary computing, and intelligent systems, instrumentation subject to critical conditions, automotive, marine and aero-space control and all other control applications, Intelligent Control System, Wiring/Wireless Sensor, Signal Control System. Sensors, Actuators and Systems Integration : Intelligent sensors and actuators, multisensor fusion, sensor array and multi-channel processing, micro/nano technology, microsensors and microactuators, instrumentation electronics, MEMS and system integration, wireless sensor, Network Sensor, Hybrid

Sensor, Distributed Sensor Networks. Signal and Image Processing : Digital signal processing theory, methods, DSP implementation, speech processing, image and multidimensional signal processing, Image analysis and processing, Image and Multimedia applications, Real-time multimedia signal processing, Computer vision, Emerging signal processing areas, Remote Sensing, Signal processing in education. Industrial Informatics: Industrial applications of neural networks, fuzzy algorithms, Neuro-Fuzzy application, bioInformatics, real-time computer control, real-time information systems, human-machine interfaces, CAD/CAM/CAT/CIM, virtual reality, industrial communications, flexible manufacturing systems, industrial automated process, Data Storage Management, Harddisk control, Supply Chain Management, Logistics applications, Power plant automation, Drives automation. Information Technology, Management of Information System : Management information systems, Information Management, Nursing information management, Information System, Information Technology and their application, Data retrieval, Data Base Management, Decision analysis methods, Information processing, Operations research, E-Business, E-Commerce, E-Government, Computer Business, Security and risk management, Medical imaging, Biotechnology, Bio-Medicine, Computer-based information systems in health care, Changing Access to Patient Information, Healthcare Management Information Technology. Communication/Computer Network, Transportation Application : On-board diagnostics, Active safety systems, Communication systems, Wireless technology, Communication application, Navigation and Guidance, Vision-based applications, Speech interface, Sensor fusion, Networking theory and technologies, Transportation information, Autonomous vehicle, Vehicle application of affective computing, Advance Computing technology and their application : Broadband and intelligent networks, Data Mining, Data fusion, Computational intelligence, Information and data security, Information indexing and retrieval, Information processing, Information systems and applications, Internet applications and performances, Knowledge based systems, Knowledge management, Software Engineering, Decision making, Mobile networks and services, Network management and services, Neural Network, Fuzzy logics, Neuro-Fuzzy, Expert approaches, Innovation Technology and Management : Innovation and product development, Emerging advances in business and its applications, Creativity in Internet management and retailing, B2B and B2C management, Electronic transceiver device for Retail Marketing Industries, Facilities planning and management, Innovative pervasive computing applications, Programming paradigms for pervasive systems, Software evolution and maintenance in pervasive systems, Middleware services and agent technologies, Adaptive, autonomic and context-aware computing, Mobile/Wireless computing systems and services in pervasive computing, Energy-efficient and green pervasive computing, Communication architectures for pervasive computing, Ad hoc networks for pervasive communications, Pervasive opportunistic communications and applications, Enabling technologies for pervasive systems (e.g., wireless BAN, PAN), Positioning and tracking technologies, Sensors and RFID in pervasive systems, Multimodal sensing and context for pervasive applications, Pervasive sensing, perception and semantic interpretation, Smart devices and intelligent environments, Trust, security and privacy issues in pervasive systems, User interfaces and interaction models, Virtual immersive communications, Wearable computers, Standards and interfaces for pervasive computing environments, Social and economic models for pervasive systems, Active and Programmable Networks, Ad Hoc & Sensor Network, Congestion and/or Flow Control, Content Distribution, Grid Networking, High-speed Network Architectures, Internet Services and Applications, Optical Networks, Mobile and Wireless Networks, Network Modeling and Simulation, Multicast, Multimedia Communications, Network Control and Management, Network Protocols, Network Performance, Network Measurement, Peer to Peer and Overlay Networks, Quality of Service and Quality of Experience, Ubiquitous Networks, Crosscutting Themes – Internet Technologies, Infrastructure, Services and Applications; Open Source Tools, Open Models and Architectures; Security, Privacy and Trust; Navigation Systems, Location Based Services; Social Networks and Online Communities; ICT Convergence, Digital Economy and Digital Divide, Neural Networks, Pattern Recognition, Computer Vision, Advanced Computing Architectures and New Programming Models, Visualization and Virtual Reality as Applied to Computational Science, Computer Architecture and Embedded Systems, Technology in Education, Theoretical Computer Science, Computing Ethics, Computing Practices & Applications

Authors are invited to submit papers through e-mail ijcsiseditor@gmail.com. Submissions must be original and should not have been published previously or be under consideration for publication while being evaluated by IJCSIS. Before submission authors should carefully read over the journal's Author Guidelines, which are located at <http://sites.google.com/site/ijcsis/authors-notes> .



© IJCSIS PUBLICATION 2016

ISSN 1947 5500

<http://sites.google.com/site/ijcsis/>